The Hong Kong Polytechnic University

Face Image Analysis and Its Applications

A thesis submitted in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

Student Name: **Xie Xudong**

Supervisor: **Dr. Kenneth K. M. Lam**

Date: **August 2005**

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 acknowledgement has been made in the text.

_____(Signed)

XIE Xudong (Name of student)

Abstract

The aim of this research is to develop efficient algorithms for facial image analysis. Our research focuses on three areas: face recognition, illumination models and compensation, and facial expression recognition. We also review some well-known face recognition techniques and the recent development of the methods for face recognition under varying illuminations and the methods for facial expression recognition.

We have proposed two methods for face recognition under various conditions: Elastic Shape-Texture Matching (ESTM) and Doubly nonlinear mapping kernel Principal Component Analysis (DKPCA). ESTM uses not only the shape information but also the texture information in comparison of two faces without establishing any precise pixel-wise correspondence. Because elastic matching is carried out within the neighborhood of each edge pixel concerned, which is robust to small, local distortions of the feature points such as facial expression variations, this method is robust to small shape variations. DKPCA is a Gabor-based method which uses the Gabor wavelets to extract facial features. Then, a doubly nonlinear mapping kernel PCA is proposed to perform feature transformation and face recognition. The proposed nonlinear mapping not only considers the statistical property of the input features, but also adopts an eigenmask to emphasize those important facial feature points. Therefore, after this mapping, the transformed features have a higher discriminating power, and the relative importance of the

features adapts to the spatial importance of the face images. This new nonlinear mapping is combined with the conventional kernel PCA for face recognition.

Lighting conditions have a serious impact on the performance of face recognition methods. Most of them will perform poorly under various conditions. Therefore, we investigate and propose two model-based methods for modeling illumination on the human face, so the effect of uneven lighting can be reduced or compensated for. Depending on the illumination model and human face model used, we model an illumination using a series of multiplicative factors and additive factors, which can be determined by the illumination model concerned and the shape of a human face. The first method can compensate for the uneven illuminations on human faces and reconstruct face images in normal lighting conditions, where a 2D face shape model is used to obtain a shape-free texture image. Instead of computing the multiplicative factors and the additive factors, the second illumination compensation method proposed in this thesis aims to reduce or even remove the effect of these factors. In this method, a local normalization technique is applied to an image, which can effectively and efficiently eliminate the effect of uneven illuminations while keeping the local statistical properties of the processed image the same as in the corresponding image under normal lighting conditions. After processing, the image under varying illumination will have similar pixel values to the corresponding image under normal lighting conditions. Then, the processed images can be used for face recognition.

We have also presented an efficient method for facial expression recognition. We first propose a representation model for facial expressions, namely spatially maximum occurrence model (SMOM), which is based on the statistical characteristics of training facial images and has a powerful representation capability. The ESTM algorithm is then used to measure the similarity between images for facial expression recognition. By combining SMOM and ESTM, the algorithm is called SMOM-ESTM and can achieve a higher recognition performance level.

To reduce the computational complexity when face recognition is applied to a large-scale database, it is necessary to filter the large database to form a smaller one that contains face images similar to the query input. Therefore, we propose an efficient indexing structure for searching a human face in a large database, which can produce a condensed database including the target image and therefore reduce the search time.

All these methods proposed in this thesis have been evaluated and compared to the existing methods. Experimental results show that our algorithms can have convincing and consistent performances.

List of Publications

The following technical papers have been published or accepted for publication based on the result generated from this work.

Journal Papers (Accepted)

1. Xudong Xie and Kin-Man Lam, "Gabor-Based Kernel PCA with Doubly Nonlinear Mapping for Face Recognition with a Single Face Image", accepted to appear in *IEEE Transactions on Image Processing*.

2. Xudong Xie and Kin-Man Lam, "An Efficient Illumination Normalization Method for Face Recognition", *Pattern Recognition Letters*, vol. 27, no. 6, pp. 609-617, 2006.

3. Xudong Xie and Kin-Man Lam, "Face Recognition under Varying Illumination Based on a 2D Face Shape Model", *Pattern Recognition*, vol. 38, no. 2, pp. 221-230, 2005.

Journal Papers (Submitted)

1. Xudong Xie and Kin-Man Lam, "Elastic Shape and Texture Matching for Human Face Recognition", revised version submitted to *IEEE Transactions on Image Processing*.

2. Xudong Xie and Kin-Man Lam, "Facial Expression Recognition based on Shape and Texture", submitted to *IEEE Transactions on Systems, Man and Cybernetics - Part A*: *Systems and Humans*.

Conference Papers (Accepted)

1. Xudong Xie and Kin-Man Lam, "Kernel PCA with Doubly Nonlinear Mapping for Face Recognition", Proceedings, the *2005 International Symposium on Intelligent Signal Processing and Communications Systems*, *ISPACS 2005*, Hong Kong, 14 – 17 Dec., 2005.

2. Xudong Xie and Kin-Man Lam, "Efficient Human Face Recognition based on Shape and Texture", Proceedings, the *Seventh International Conference on Signal Processing*, *ICSP 2004*, Beijing, China, 31 Aug. - 4 Sept., 2004.

3. Xudong Xie and Kin-Man Lam, "An Efficient Illumination Compensation Scheme for Face Recognition", Proceedings, the *Eighth International Conference on Control, Automation, Robotics and Vision, ICARCV 2004*, Kunming, China, 6 - 9 Dec., 2004.

4. Xudong Xie and Kin-Man Lam, "An Efficient Method for Face Recognition under Varying Illumination", Proceedings, *IEEE International Symposium on Circuits and Systems*, *ISCAS 2005*, Kobe, Japan, 23 - 26 May, 2005.

5. Xudong Xie and Kin-Man Lam, "An Efficient Method for Facial Expression Recognition", Proceedings, the *Visual Communications and Image Processing 2005*, *VCIP 2005*, Beijing, China, 12 - 15 July, 2005.

6. Kwan-Ho Lin, Kin-Man Lam, Xudong Xie and Wan-Chi Siu, "An Efficient Human Face Indexing Scheme Using Eigenfaces", Proceedings, *IEEE International Conference on Neural Networks and Signal Processing*, Nanjing, China, 14 - 17 Dec., 2003.

Acknowledgements

I would like to take this opportunity to express my sincere gratitude to my Supervisor, Dr. K. M. Lam, from the Department of Electronic and Information Engineering of The Hong Kong Polytechnic University, for his patient and guidance throughout my research studies. My Ph.D. studies would have never been completed without his supervision. His immense enthusiasm for research is greatly appreciated. His professional advice and plentiful experience help me to further improve myself.

I am also thankful to all the members of the DSP Research Laboratory, especially for Professor W. C. Siu, Dr. Danghui Liu, Dr. Cheung-Ming Lai, Mr. Yong Xia and Mr. K. W. Wong for their supportive and constructive comments that will never be forgotten. Their expert knowledge and friendly encouragement have helped me to overcome a lot of difficulties during my research studies. I would also like to thank the Centre for Multimedia Signal Processing of the Department of Electronic and Information Engineering for generous support over the past three years.

Last but not least, without the patience and forbearance of my family, the preparation of this research work would have been impossible. I appreciate their constant and continuous support and understanding.

List of Abbreviations

AFA	Automatic Face Analysis
AHE	Adaptive Histogram Equalization
AP	Action Parameter
AU	Action Unit
BHE	Block-based Histogram Equalization
CDA	Clustering-based Discriminant Analysis
DKPCA	Doubly nonlinear mapping Kernel PCA
DLA	Dynamic Link Architecture
EAT	Elsevier Advanced Technology
EGM	Elastic Graph Matching
ESTM	Elastic Shape-Texture Matching
FACS	Facial Action Coding System
FFT	Fast Fourier Transform
FPP	Fractional Power Polynomial
GW	Gabor Wavelet
HE	Histogram Equalization
HMM	Hidden Markov Model
IC	Independent Component
ICA	Independent Component Analysis
IFFT	Inverse Fast Fourier Transform
IM	Illumination Map
KFDA	Kernel Fisher Discriminant Analysis

KPCA	Kerenl Principal Component Analysis
LDA	Linear Discriminant Analysis
LEM	Line Edge Map
LN	Local Normalization
LPP	Locality Preserving Projections
M2HD	Doubly Modified Hausforff Distance
MHD	Modified Hausdorff Distance
ODV	Optimal Discriminant Vector
OSH	Optimal Separating Hyperplane
PCA	Principal Component Analysis
PDBNN	Probabilistic Decision-Based Neural Network
PDF	Probability Density Function
RBF	Radial Basis Function
SMOM	Spatially Maximum Occurrence Model
SVM	Support Vector Machine

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

The objective of this chapter is to introduce the general concepts of face image analysis, including the characteristics of human faces, and some applications of the face-based techniques. We will also address the originality and the organization of this thesis.

1.1 Motivation

The human face plays an important role in personal communication. Firstly, due to the uniqueness of the face of a person, it can be considered as the personal ID, which can comprise part of many applications [1-5], such as criminal identification, credit card verification, security system, scene surveillance, entertainments, etc. In fact, humans have used faces to recognize each other for thousands of years [6, 7]. Secondly, people mainly express their emotions through different facial expressions and tones of voice. Just as a mid-16th century proverb says, “the eyes are the window of the soul”, so the change of facial expression intuitively reflects the latent emotion. Furthermore, social psychology research has shown that facial expressions convey messages more powerful than the spoken words in meaningful conversations [8]. Finally, the face assists in a number of cognitive tasks in speach recognition; for example, the shape and motion of lips can contribute greatly to speech comprehension in a noisy environment. Therefore, the face can be considered the personal communication center.

Compared with other biometric characteristics, such as fingerprints, hand geometry, iris, retina, etc., face-based applications are more user-friendly and non-intrusive. That is, the system has the ability to measure the characteristic, i.e. the face image, of an individual without contact. In addition, only very little cooperation or participation from the users is required. This property is very useful for some security applications [3-5]. Besides the above mentioned applications, face analysis techniques can also be applied to natural human-computer interface systems [9], such as virtual reality, computer games, robotic dogs, and so on. In 2002, the Elsevier Advanced Technologies' (EAT) report [10] quoted the facial recognition market as being US$32.9 million, and by 2006 this amount will have grown to US$242.7 million. We can see that face analysis techniques can be used in a myriad of commercial and law enforcement applications, which are potentially huge markets.

Over the past few years, face image analysis has attracted researchers from disciplines such as image processing, pattern recognition, neural networks, computer vision, computer graphics, and psychology. The different applications have developed different techniques, such as face detection [11-15], face recognition [1, 2, 5, 16, 17], face tracking [18-20], facial expression recognition [5, 21-25], gender determination [26-28], age classification [29, 30], aging simulation [29, 31], face synthesizing [32-34] and 3D face analysis [35-38]. In this thesis, we mainly consider the techniques for face recognition and facial expression recognition. Finally, we also investigate and devise a new indexing structure for searching for a particular face image from a large face database.

1.2 Statements of Originality

The following contributions reported in this thesis are claimed to be original.

1. A new elastic shape-texture matching method, namely ESTM, for human face recognition is derived. In our approach, both the shape and texture information are used to compare two faces without establishing any precise pixel-wise correspondence. Combining the shape and texture features together, a shape-texture Hausdorff distance is devised to compute the similarity between two face images.
2. A novel Gabor-based kernel Principal Component Analysis (PCA) with doubly nonlinear mapping is proposed for human face recognition. In this method, the Gabor wavelets are used to extract facial features, then a doubly nonlinear mapping kernel PCA is proposed to perform feature transformation and face recognition.
3. A simple yet effective local contrast enhancement method, namely block-based histogram equalization (BHE), is proposed to estimate the category of the light source of an input face image.
4. A novel illumination compensation algorithm, which can compensate for the uneven illuminations on human faces and reconstruct face images in normal lighting conditions, is proposed. Based on the light category identified, a corresponding lighting compensation model is used to reconstruct an image that will visually be under normal illumination. In order to eliminate the influence of uneven illumination while retaining the shape information about a human face, a 2D face shape model is used.

5. An efficient representation method insensitive to varying illumination is presented for human face recognition. This method applies a local normalization technique to an image, which can effectively and efficiently eliminate the effect of uneven illuminations while keeping the local statistical properties of the processed image the same as in the corresponding image under normal lighting condition.
6. A representation model for facial expressions, namely spatially maximum occurrence model (SMOM), which is based on the statistical characteristics of training facial images and has a powerful representation capability, is proposed.
7. An efficient method for human facial expression recognition is devised. Combining ESTM algorithm with SMOM, a new method called SMOM-ESTM is used for facial expression recognition.
8. An efficient indexing structure for searching a human face in a large database is also proposed. This method will form a small database, namely a *condensed database*, for face recognition, instead of considering the original large database.

1.3 Outline of the Thesis

This thesis is organized into nine chapters and each chapter is outlined as follows.

Chapter 2 describes the principles of face recognition and facial expression recognition. We will briefly review some well-known face recognition techniques, such as Principal Component Analysis (PCA) [39-41], Linear Discriminant Analysis (LDA) [42], Independent Component Analysis (ICA) [43-46], Kerenl Principal Component Analysis (KPCA) [47-50], Hausdorff distance [51, 52] and Gabor

wavelets [44, 53, 54]. These methods are related to our methods proposed in this thesis, which will be described in the following chapters. We will also review the recent development of the face recognition methods for varying illuminations and of the methods for facial expression recognition. We will also compare in this thesis our proposed algorithms to some of the existing ones.

In Chapter 3, we introduce a novel elastic shape-texture matching method, namely ESTM, for human face recognition. In our approach, both the shape and texture information are used to compare two faces without establishing any precise pixel-wise correspondence. The edge map is used to represent the shape of an image and is allowed to act as an elastic graph when performing matching, and the texture information is characterized by both the Gabor representations and the gradient direction of each pixel. Combining these features, a shape-texture Hausdorff distance is devised to compute the similarity between two face images. The elastic matching is carried out within the neighborhood of each edge pixel concerned, which is robust to small, local distortions of the feature points, such as facial expression variations. Due to the fact that the edge map, Gabor representations and the direction of image gradient can all alleviate the effect of illumination, ESTM is therefore robust to lighting condition variations.

Chapter 4 presents a novel Gabor-based kernel Principal Component Analysis (KPCA) with doubly nonlinear mapping for human face recognition. In our approach, the Gabor wavelets are used to extract facial features, then a doubly nonlinear mapping kernel PCA is proposed to perform feature transformation and face recognition. The conventional kernel PCA nonlinearly maps an input image

into a high-dimensional feature space in order to make the mapped features linearly separable. However, this method does not consider the structure of the manifold on which the face images possibly reside, and it is difficult to determine which nonlinear mapping is more effective for face recognition. In this chapter, a new method of nonlinear mapping, which is performed in the original feature space, is defined. The proposed nonlinear mapping not only considers the statistical property of the input features, but also adopts an eigenmask [55, 56] to emphasize those important facial feature points. Therefore, after this mapping, the transformed features have a higher discriminating power, and the relative importance of the features adapts to the spatial importance of the face images. This new nonlinear mapping is combined with the conventional kernel PCA to be called 'doubly' nonlinear mapping kernel PCA (DKPCA).

In Chapter 5, we propose a novel illumination compensation algorithm, which can compensate for the uneven illuminations on human faces and reconstruct face images in normal lighting conditions. According to the illumination model and human face model used, the effect of uneven illumination can be modeled as a series of multiplicative factors and additive factors, which can be determined by the illumination model concerned and the shape of a human face. To eliminate the influence of shape on different faces, a 2D face shape model is used to obtain a shape-free texture image. For an identified illumination category, the effect of a particular uneven lighting, i.e. a particular multiplicative factor and additive factor, can be computed using a set of training images, and are used for reconstructing an

image that will visually be under normal illumination. Then, these images can be used for face recognition.

In Chapter 6, an efficient representation method insensitive to varying illumination is presented for human face recognition. Instead of computing the multiplicative factors and the additive factors, which are used to model the uneven illuminations for face recognition as described in Chapter 5, we aim to reduce or even remove the effect of these factors. In our method, a local normalization technique is applied to an image, which can effectively and efficiently eliminate the effect of uneven illuminations while keeping the local statistical properties of the processed image the same as in the corresponding image under normal lighting condition. After processing, the image under varying illumination will have similar pixel values to the corresponding image that is under normal lighting condition.

Chapter 7 presents an efficient method for human facial expression recognition. We first propose a representation model for facial expressions, namely spatially maximum occurrence model (SMOM), which is based on the statistical characteristics of training facial images and has a powerful representation capability. The ESTM algorithm is then used to measure the similarity between images for facial expression recognition. By combining SMOM and ESTM, the algorithm is called SMOM-ESTM and can achieve a higher recognition performance level.

In Chapter 8, an efficient indexing structure for searching a human face in a large database is proposed. In our method, a set of eigenfaces is computed based on the faces in the database. Each face in the database is then ranked according to its

projection onto each of the eigenfaces. A query input will be ranked similarly, and the corresponding nearest faces in the ranked position with respect to each of the eigenfaces are selected from the database. These selected faces will then form a small database, namely a condensed database, for face recognition, instead of considering the original large database.

Finally, we give the conclusions of our work in Chapter 9, where some suggestions for further development are also provided.

Chapter 2. Literature Review

In this chapter, we introduce the general concepts of face recognition and facial expression recognition. We briefly review some well-known face recognition techniques related to the methods that we propose in this thesis. We also review the recent development of the methods for face recognition under varying illuminations and the methods for facial expression recognition, some of which will be compared to our proposed methods in the chapters that follow.

2.1 Review of Face Recognition

2.1.1 Problem Statement

The face recognition problem is to automatically recognize the identity of a person from a new image by comparing it to human facial images annotated with identity in a stored database. In other words, we need to find the pictures of the same person as the input image that are in a large facial image database. Here we suppose that the location of a face in an image is known, so that we only need to consider the similarity between different facial images. Chellappa et al. [2] gave a more general statement for face recognition, which includes face detection from a scene, feature extraction from the face region, and feature matching for comparison. For a real face recognition application, we actually should perform these procedures, which starts with face detection. However, considering the different characteristics of face detection and face recognition, it is wise to divide them into

two stages and treat them separately. In fact, there have been many methods proposed for face detection [11-15, 57-60].

In real applications, face recognition techniques use various source formats ranging from static, controlled format photographs to uncontrolled video sequences, all of which have been produced in different conditions. Therefore, a practical face recognition technique needs to be robust to the image variations caused by different factors, such as:

1. **Pose:** The images of a face vary due to the relative camera-face pose, and some facial features such as the eyes or the nose may become partially or wholly occluded,
2. **Presence or absence of structural components:** Facial features such beards, mustaches, and glasses may or may not be present, and there is a great deal of variability among these components including shape, color, and size,
3. **Facial expression:** The appearance of faces is directly affected by a person's facial expression,
4. **Occlusion:** A face may be partially occluded by other objects [61]. In an image with a group of people, some faces may partially occlude each other,
5. **Image orientation:** Face images vary for different rotations about the camera's optical axis, and
6. **Image conditions:** When an image is formed, factors such as lighting and camera characteristics (sensor response, lenses) affect the appearance of the faces in the image.

In fact, there are two kinds of classification problems in face image analysis applications. The first is how to recognize a human face under the above mentioned variations, and the second is how to estimate the characteristics of a person, such as the age [29, 30], gender [26-28], hairstyle [62], expression [5, 21-25] and pose [63-65], or the situation of the picture, e.g. the illumination [66-68]. These two problems are not isolated. In most cases, if we know who the person is, we can also judge his/her characteristics, such as age, gender, ethnic origin, etc. Similarly, if we have some information about the target, the corresponding compensation operation can be performed, or the search range can be greatly reduced, which accordingly results in a more accurate recognition result.

In this thesis, Chapters 3 and 4 will describe our proposed methods for face recognition under various conditions. The methods presented in Chapters 5 and 6 address the problem of face recognition under varying illuminations, which can be considered a special application for the first-class classification problem. In fact, in the method described in Chapter 5, a block-based histogram equalization (BHE) method is proposed to estimate the illumination category, which belongs to the second-class classification problem. Chapter 7 presents a method for facial expression recognition, which is a classical second-class application. Finally, the database condensing technique proposed in Chapter 8 is not a direct recognition application, but this method can narrow the searching range when face matching is performed.

2.1.2 History and Development of Face Recognition

It is well known that humans have used their faces to recognize each other for thousands of years [6, 7]. The earliest work on face recognition can be traced back at least to the 1950s in psychology [69] and to the 1960s in the engineering literature [70]. In fact, Darwin did some work on facial profile-based biometrics in 1888 [71]. However, research on automatic machine recognition of faces really started in the 1970s [72] after the seminal work of Kanade [73].

Over the past 30 years, psychophysicists and neuroscientists have been concerned with issues such as whether face perception is a dedicated process and whether it is done holistically or by local feature analysis [74, 75]. These findings have been combined with various techniques, such as image processing, pattern recognition, neural networks, computer vision, computer graphics, etc., to develop a sequence of algorithms and systems for machine recognition of human faces. In the last decade in particular, many significant advances have taken place. In the following section, some of the existing face recognition algorithms will be introduced and discussed.

2.1.2.1 Linear Subspace Analysis

Linear subspace analysis, which considers a feature space as a linear combination of a set of bases, has been widely used in face recognition applications. This is mainly due to its effectiveness and computational efficiency for feature extraction and representation. Different criteria will produce different bases and, consequently, the transformed subspace will also have different properties. Principal Component Analysis (PCA) [40, 41, 76], which is widely used

for face recognition and face reconstruction, decomposes an input image as a combination of a sequence of basis images, namely eigenfaces, and therefore has a low computational complexity and high representation ability. In 1997, Linear Discriminant Analysis (LDA) [42] is proposed, which not only maximizes the between-class scatters of different subjects, but also minimizes the within-class scatters of the same person when performing feature transformation. Therefore, LDA can preserve the discriminating information and is suitable for recognition. Because only the second-order dependencies in the PCA coefficients are eliminated, PCA cannot capture even the simplest invariance unless this information is explicitly provided in the training data [77]. Independent Component Analysis (ICA), which was proposed in 2002 [43], can be considered a generalization of PCA, and aims to find some independent bases by methods sensitive to high-order statistics. As opposed to PCA, 2DPCA [78] is based on 2D image matrices rather than 1D vectors so the image matrix does not need to be transformed into a vector prior to feature extraction. Instead, an image covariance matrix is constructed directly using the original image matrices, and its eigenvectors are derived for image feature extraction. Locality Preserving Projections (LPP) [79] obtains a face subspace that best detects the essential face manifold structure, and preserves the local information of the image space. When the proper dimension of the subspace is selected, the recognition rates using LPP are better than those using PCA or LDA, based on different databases.

2.1.2.2 Kernel-Based Methods

With the Cover's theorem, nonlinearly separable patterns in an input space will become linearly separable with a high probability if the input space is transformed nonlinearly to a high-dimensional feature space [36]. We can therefore map an input image into a high-dimensional feature space, so that linear discriminant methods can then be employed for face recognition. This mapping is usually realized via a kernel function [80] and, according to the methods used for recognition in the high-dimensional feature space, we have a set of kernel-based methods, such as the Kernel PCA (KPCA) [48-50], or the Kernel Fisher discriminant analysis (KFDA) [80-84]. KPCA and KFDA are linear in the high-dimensional feature space, but nonlinear in the low-dimensional image space. In other words, these methods can discover the nonlinear structure of the face images, and encode higher order statistics [50].

Support vector machine (SVM) [85-87], a pattern classification algorithm developed by V. Vapnik and his co-operators [87, 88], finds the hyperplane that separates the largest possible fraction of points of the same class on the same side, while maximizing the distance from either class to the hyperplane, for a two-class classification problem. According to Vapnik [89], this hyperplane is called Optimal Separating Hyperplane (OSH), which minimizes the risk of misclassifying not only the examples in the training set but also the unseen examples of the test set. The attractiveness of using neural network could be due to its nonlinearity in the network.

2.1.2.3 Neural Network

One of the first artificial neural network techniques used for face recognition is a single layer adaptive network called WISARD, which contains a separate network for each stored individual [90]. However, when the number of persons increases, the computing expense will become more demanding. The probabilistic decision-based neural network (PDBNN) [91] is effectively applied to face detection and recognition. PDBNN has inherited the modular structure from its predecessor described in [92]. PDBNN-based identification systems have the merits of both neural networks and statistical approaches, and their distributed computing principle is relatively easy to implement on parallel computers.

A radial basis function (RBF) neural classifier is used to cope with small training sets of high dimension, which is a problem frequently encountered in face recognition in 2002 [93]. In order to avoid overfitting and reduce the computational burden, face features are first extracted by the PCA method. Then, the resulting features are further processed by the LDA technique to acquire lower-dimensional discriminant patterns. A paradigm is proposed whereby data information is encapsulated in determining the structure and initial parameters of the RBF neural classifier before learning takes place. A hybrid learning algorithm is used to train the RBF neural networks so that the dimension of the search space is drastically reduced in the gradient paradigm.

2.1.2.4 Graph Matching

In [54], a dynamic link architecture (DLA) for distortion invariant object recognition is presented. The DLA first computes the Gabor jets of the face images,

and then elastic graph matching (EGM) is used to compare their resulting image decompositions. Duc et al. [94] introduced an automatic weighting for the nodes of the elastic graph according to their significance, and also explored the significance of the elastic deformation for an application of face-based person authentication. Kotropoulos et al. [95] has proposed a morphological dynamic link architecture which adopts discriminatory power coefficients to weigh the matching error at each grid node. In general, these methods can preserve some texture features and local geometry information [96], and therefore are superior to other face recognition techniques in terms of rotation invariant; however, the matching process is computationally expensive.

2.1.2.5 Hidden Markov Models

Stochastic modeling of non-stationary vector time series based Hidden Markov model (HMM) has been very successful for speech applications. Samaria et al. [97] first applied this method to human face recognition. Samaria et al. [98] proposed to model human faces with a vertical top-to-bottom 1D HMM structure composed of superstates. Each superstate contains a horizontal left-to-right 1D Markov chain. In [99], a similar 1D HMM, which uses 2D-DCT coefficients as the feature vectors of the HMM, is proposed. Due to the compression properties of the DCT, the size of the observation vector is reduced, while preserving the same recognition rate. The embedded HMMs [100] models the two-dimensional data better than the one-dimensional HMM and is computationally less complex than the two-dimensional HMM. This model is appropriate for face images since it exploits an important facial characteristic: frontal faces preserve the same structure

of “super states” from top to bottom, and also the same left-to-right structure of “states” inside each of these “super states”. Embedded Bayesian network [101], a generalized framework of embedded HMM, is defined recursively as a hierarchical structure where the “parent” node is a Bayesian network that conditions the embedded Bayesian networks or the observation sequence that describes the nodes of the “child” layer. Embedded Bayesian network shows a significant complexity reduction. 2D HMM [102] builds on an assumption of conditional independence in the relationship between adjacent blocks. This allows the state transition to be separated into vertical and horizontal state transitions. This separation of state transitions brings the complexity of the hidden layer of the proposed model from the order of (N^3T) to the order of ($2N^2T$), where N is the number of the states in the model and T is the total number of observation blocks in the image. The system is tested on the facial database of AT&T Laboratories Cambridge and the more complex facial database of the Georgia Institute of Technology where recognition rates up to 100 percent and 92.8 percent have been achieved, respectively, with relatively low complexity. In [103], HMM was used in the temporal domain to perform face recognition in video signals, where each frame in the video sequence is considered as an observation.

2.1.2.6 Geometrical Feature Matching and Template Matching

Geometrical Feature Matching techniques are based on the computation of a set of geometrical features from the picture of a face. Bruneli et al. [104] automatically extracted geometrical features, such as nose width and length, mouth position, and chin shape, and used a Bayes classifier to face recognition. Manjunath

et al. [105] used Gabor wavelet decomposition to detect feature points for each face image which greatly reduced the storage requirement for the database. Tamura et al. [106] found that face recognition based on geometrical feature matching is possible for face images at resolution as low as 8×6 pixels when single facial features are hardly revealed. In summary, geometrical feature matching based on precisely measured distances between features may be most useful for finding possible matches in a large database such as a mug shot album. However, it will be dependent on the accuracy of the feature location algorithms. Current automated face feature location algorithms do not provide a high degree of accuracy and require considerable computational time.

In template matching methods, several standard patterns for a face are stored to describe the face as a whole or the facial features separately. The correlations between an input image and the stored patterns are computed for detection. The templates are allowed to translate, scale, and rotate. Segments obtained from the curvature discontinuities of the head outline can be used as templates. A simple version of template matching is that a test image represented as a two-dimensional array of intensity values is compared using a suitable metric, such as the Euclidean distance, with a single template representing the whole face. There are several other more sophisticated methods based on template matching for face recognition [107, 108].

Line Edge Map (LEM) [109] approach, which extracts lines from a face edge map as features, can be considered as a combination of template matching and geometrical feature matching. LEM integrates the structural information with

spatial information of a face image by grouping pixels of face edge map to line segments. After thinning the edge map, a polygonal line fitting process [110] is applied to generate the LEM of a face. The LEM representation, which records only the end points of line segments on curves, further reduces the storage requirement. LEM is also expected to be less sensitive to illumination changes due to the fact that it is an intermediate-level image representation derived from low-level edge map representation. Therefore, the LEM approach not only possesses the advantages of feature-based approaches, such as invariant to illumination and low memory requirement, but also has the advantage of high recognition performance of template matching. Comparing with LEM, a more reliable method, Elastic Shape-Texture Matching (ESTM) [111], which is also based on the combination of template matching and geometrical feature matching, is proposed. In ESTM, the edge map is used to represent the shape of an image and is allowed to act as an elastic graph when performing matching, and the texture information is characterized by both the Gabor representations and the gradient direction of each pixel. Combining these features, a shape-texture Hausdorff distance is devised to compute the similarity between two face images.

2.1.3 Some Related Methods

In this section, we will briefly introduce some techniques that are related to our approaches proposed in the later chapters.

2.1.3.1 Principal Component Analysis

PCA is a classical method that has been widely used for human face representation and recognition. The major idea of PCA is to decompose a data space

into a linear combination of a small collection of bases, which are pairwise orthogonal and which capture the directions of maximum variance in the training set. Suppose there are a set of centered *N*-dimensional training samples $\mathbf{Y}_i$, $i = 1, 2, \cdots, M$, such that $\mathbf{Y}_i \in R^N$ and $\sum_{i=1}^{M} \mathbf{Y}_i = \vec{0}$. The covariance matrix of the input can be estimated as follows:

$$\Sigma = \frac{1}{M} \sum_{i=1}^{M} \mathbf{Y}_i \mathbf{Y}_i^T \ . \tag{2.1}$$

The PCA leads to solve the following eigenvector problem:

$$\lambda \mathbf{v} = \Sigma \mathbf{v} \ , \tag{2.2}$$

where $\mathbf{v}$ are the eigenvectors of Σ, and λ are the corresponding eigenvalues. These eigenvectors are ranked in a descending order according to the magnitudes of their eigenvalues, and the first *L* (generally, $L < N$) eigenvectors are selected as the bases, which are commonly called eigenfaces. These eigenfaces with large eigenvalues represent the global, rough structure of the training images, while the eigenfaces with small eigenvalues are mainly determined by the local, detailed components. Therefore, after projecting onto the eigenspace, the dimension of the input is reduced while the main components are maintained. For face recognition, when the testing images have variations caused by local deformation, such as different facial expressions [112], PCA can alleviate this effect. However, when the variations are caused by global components such as lighting or perspective variations, the performance of PCA will be greatly degraded [67].

2.1.3.2 Independent Component Analysis

PCA can remove the pair-wise linear dependencies between pixels in an image, but high-order dependencies still exist in the joint distribution of the PCA coefficients. ICA [43-46] can be considered a generalization of PCA, which can find some independent bases, namely Independent Components (ICs), by methods sensitive to high-order statistics. Suppose **s** is the vector of unknown source image, and **Y** is the vector of observed mixtures. If **A** is the unknown mixing matrix, then the mixing process is shown as

$$\mathbf{Y} = \mathbf{As}. \tag{2.3}$$

The goal of ICA is find the separating matrix **W** such that

$$\mathbf{s} = \mathbf{WY}. \tag{2.4}$$

However, there is no closed form expression to find **W**. Instead, many iterative algorithms are used to approximate **W** in order to optimize independence of **Y**. According to [43], there are two types of implementation frameworks for ICA in the image recognition task. Framework I treats images as random variables and pixels as observations; while Framework II coins pixels as random variables and images as observations. In framework I, the basis vectors obtained are approximately independent, but the coefficients representing each image are not necessarily independent. On the other hand, framework II finds a representation in which all the coefficients are statistically independent. Therefore, framework I and II can be interpreted as local features and global texture features extractor, respectively. ICA architecture I is used for localized tasks, and ICA architecture II for holistic tasks. There are different algorithms implemented for different ICA architecture, e.g. the

InfoMax [113] approach for the ICA architecture I, and the FastICA [114] algorithm for architecture II.

Although the ICs are independent to each other, while the eigenfaces are uncorrelated to each other, we cannot argue that ICA always performs better than PCA for face recognition. Bartlett, *et al*. [115, 116], Liu and Wechsler [117], and Yuen and Lai [118] claim that ICA outperforms PCA for face recognition, while Baek *et al*. [119] claim that PCA outperforms ICA and Moghaddam [120] claims that there is no statistical difference in performance between the two. The experimental results in [46] show that comparisons between PCA and ICA are complex, because differences in tasks, architectures, ICA algorithms, and distance metrics must be taken into account.

2.1.3.3 Linear Discriminant Analysis

Let $\mathbf{Y}_{ij}$ be an N-dimensional vector representing the j^{th} image of the i^{th} person, K the number of distinct persons in a database, and M the number of images of each person. The within-class scatter matrix $\mathbf{S}_w$ and the between-class scatter matrix $\mathbf{S}_b$ can be written as follows:

$$\mathbf{S}_w = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{1}{M}\sum_{j=1}^{M}\left(\mathbf{Y}_{ij}-\bar{\boldsymbol{\mu}}_i\right)\left(\mathbf{Y}_{ij}-\bar{\boldsymbol{\mu}}_i\right)^T\right), \tag{2.5}$$

$$\mathbf{S}_b = \frac{1}{K}\sum_{i=1}^{K}\left(\bar{\boldsymbol{\mu}}_i-\boldsymbol{\mu}\right)\left(\bar{\boldsymbol{\mu}}_i-\boldsymbol{\mu}\right)^T, \tag{2.6}$$

where $\bar{\boldsymbol{\mu}}_i$ is the mean of the i^{th} class and $\boldsymbol{\mu}$ is the mean of all the classes. The optimal discriminant vectors $\mathbf{V}$ are computed by maximizing the following criterion:

$$J\left(\mathbf{V}\right) = \frac{\mathbf{V}^T\mathbf{S}_b\mathbf{V}}{\mathbf{V}^T\mathbf{S}_w\mathbf{V}}. \tag{2.7}$$

Then, a generalized eigenvalue problem can be solved as follows:

$$\mathbf{S}_b\mathbf{V}_i = \lambda_i\mathbf{S}_w\mathbf{V}_i, \tag{2.8}$$

where $\mathbf{V}_i$ represents an optimal vector for the criterion $J(\mathbf{V})$ and λ_i is a scalar ($i = 1, 2, ...$). If $\mathbf{S}_w$ is a nonsingular matrix, the optimal discriminant vectors (ODVs) can be solved with the following equation:

$$\left(\mathbf{S}_w^{-1}\mathbf{S}_b\right)\mathbf{V}_i = \lambda_i\mathbf{V}_i, \tag{2.9}$$

In other words, $\mathbf{V}_i$ is an eigenvector of $\mathbf{S}_w^{-1}\mathbf{S}_b$ and λ_i is the corresponding eigenvalue. The computation of the above eigenvalue problem might be unstable because the matrix $\mathbf{S}_w^{-1}\mathbf{S}_b$ may not be symmetric due to limited precision in number representation. More importantly, $\mathbf{S}_w$ is usually a singular matrix due to the small sample size. Several algorithms [17, 121-125] have been proposed to solve this problem. However, the work in [14] shows that, when the training data set is small, PCA can outperform LDA, and also that PCA is less sensitive to different training data sets.

2.1.3.4 Kernel PCA

With the Cover's theorem, nonlinearly separable patterns in an input space will become linearly separable with high probability if the input space is transformed nonlinearly into a high-dimensional feature space. We can therefore map an input variable into a high-dimensional feature space, and then perform PCA. For a given nonlinear mapping Φ, the input data space R^N can be mapped into a potentially much higher dimensional feature space F:

$$\begin{aligned} \Phi: \quad & R^N \to F, \\ & \mathbf{Y} \to \Phi(\mathbf{Y}). \end{aligned} \tag{2.10}$$

Performing PCA in the high-dimensional feature space can obtain high-order statistics of the input variables; that is also the initial motivation of the KPCA. However, it is difficult to directly compute both the covariance matrix and its corresponding eigenvectors and eigenvalues in the high-dimensional feature space. It is computationally intensive to compute the dot products of vectors with a high dimension. Fortunately, kernel tricks can be employed to avoid this difficulty, which compute the dot products in the original low-dimensional input space by means of a kernel function [47, 48]:

$$k\left(\mathbf{Y}_i, \mathbf{Y}_j\right) = \left(\Phi\left(\mathbf{Y}_i\right) \cdot \Phi\left(\mathbf{Y}_j\right)\right). \tag{2.11}$$

Define an $M \times M$ Gram matrix **R**, where M is the number of training images used, and the elements of **R** can be determined by virtue of the kernel function:

$$R_{ij} = \Phi\left(\mathbf{Y}_i\right)^T \Phi\left(\mathbf{Y}_j\right) = \left(\Phi\left(\mathbf{Y}_i\right) \cdot \Phi\left(\mathbf{Y}_j\right)\right) = k\left(\mathbf{Y}_i, \mathbf{Y}_j\right), \tag{2.12}$$

The orthonormal eigenvectors $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m$ of **R** corresponding to the m largest positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ are computed. Then, the corresponding eigenvectors $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_m$ for the KPCA can be derived as follows [48, 84]:

$$\boldsymbol{\beta}_j = \frac{1}{\sqrt{\lambda_j}} \mathbf{Q}\boldsymbol{\gamma}_j, \qquad j = 1, \cdots, m., \tag{2.13}$$

where $\mathbf{Q} = \left[\Phi\left(\mathbf{Y}_1\right), \cdots, \Phi\left(\mathbf{Y}_M\right)\right]$ is the mapped data matrix in the high-dimensional feature space. For a mapped test sample $\Phi\left(\mathbf{Y}\right)$, it should be projected onto the eigenvector system $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_m$, and the projection vector of $\Phi\left(\mathbf{Y}\right)$, $\mathbf{w} = \left(w_1, w_2, \cdots w_m\right)^T$, in the transformed subspace is computed by

$$\mathbf{w} = \mathbf{P}^T \Phi\left(\mathbf{Y}\right), \qquad \text{where } \mathbf{P} = \left(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots \boldsymbol{\beta}_m\right). \tag{2.14}$$

Specifically, the j^{th} component w_j is given as follows:

$$\begin{aligned} w_j &= \boldsymbol{\beta}_j^T \Phi(\mathbf{Y}) = \frac{1}{\sqrt{\lambda_j}} \boldsymbol{\gamma}_j^T \mathbf{Q}^T \Phi(\mathbf{Y}) \\ &= \frac{1}{\sqrt{\lambda_j}} \boldsymbol{\gamma}_j^T \left[\Phi(\mathbf{Y}_1), \cdots, \Phi(\mathbf{Y}_M) \right]^T \Phi(\mathbf{Y}) \\ &= \frac{1}{\sqrt{\lambda_j}} \boldsymbol{\gamma}_j^T \left[\Phi(\mathbf{Y}_1)^T \Phi(\mathbf{Y}), \Phi(\mathbf{Y}_2)^T \Phi(\mathbf{Y}), \cdots, \Phi(\mathbf{Y}_M)^T \Phi(\mathbf{Y}) \right] \quad (2.15) \\ &= \frac{1}{\sqrt{\lambda_j}} \boldsymbol{\gamma}_j^T \left[\Phi(\mathbf{Y}_1) \cdot \Phi(\mathbf{Y}), \Phi(\mathbf{Y}_2) \cdot \Phi(\mathbf{Y}), \cdots, \Phi(\mathbf{Y}_M) \cdot \Phi(\mathbf{Y}) \right] \\ &= \frac{1}{\sqrt{\lambda_j}} \boldsymbol{\gamma}_j^T \left[k(\mathbf{Y}_1, \mathbf{Y}), k(\mathbf{Y}_2, \mathbf{Y}), \cdots, k(\mathbf{Y}_M, \mathbf{Y}) \right], \; j = 1, \cdots, m. \end{aligned}$$

Therefore, from (2.12) and (2.15), we can see that the explicit mapping process is not required, and that all the procedures are performed in the low-dimensional input space instead of the high-dimensional feature space.

In a practical face recognition application, three classes of kernel functions have been widely used, which are the polynomial kernels, Gaussian kernels, and sigmoid kernels, [47], respectively:

Polynomial kernel: $k(\mathbf{Y}_i, \mathbf{Y}_j) = (\mathbf{Y}_i \cdot \mathbf{Y}_j)^d$, (2.16)

Gaussian kernel: $k(\mathbf{Y}_i, \mathbf{Y}_j) = \exp\left(-\frac{\left\| \mathbf{Y}_i - \mathbf{Y}_j \right\|^2}{2\sigma^2} \right)$, and (2.17)

Sigmoid kernel: $k(\mathbf{Y}_i, \mathbf{Y}_j) = \tanh\left(\kappa(\mathbf{Y}_i \cdot \mathbf{Y}_j) + \vartheta \right)$, (2.18)

where d $\in$ N, $\sigma > 0$, $\kappa > 0$, and $\vartheta < 0$. In [50], the polynomial kernels are extended to include fractional power polynomial (FPP) models, i.e. $0 < d < 1$, where a more reliable performance can be achieved.

2.1.3.5 Gabor Wavelets

The Gabor wavelets, whose kernels are similar to the response of the two-dimensional receptive field profiles of the mammalian simple cortical cell [53], exhibit the desirable characteristics of capturing salient visual properties such as spatial localization, orientation selectivity, and spatial frequency [44]. The Gabor wavelets can effectively abstract local and discriminating features, which are useful for texture detection [127] and face recognition [54, 128, 129].

In the spatial domain, a Gabor wavelet is a complex exponential modulated by a Gaussian function, which is defined as follows [53, 54, 130]:

$$\psi_{\omega,\theta}(u,v)=\frac{1}{2\pi\sigma^2}e^{-\left(\frac{(u\cos\theta+v\sin\theta)^2+(-u\sin\theta+v\cos\theta)^2}{2\sigma^2}\right)}\cdot\left[e^{i(\omega u\cos\theta+\omega v\sin\theta)}-e^{-\frac{\omega^2\sigma^2}{2}}\right], \quad (2.19)$$

where u, v denote the pixel position in the spatial domain, ω is the radial center frequency of the complex exponential, θ is the orientation of the Gabor wavelet, and σ is the standard deviation of the Gaussian function. The value of σ can be derived as follows [130]:

$$\sigma=\kappa/\omega, \quad (2.20)$$

where $\kappa=\sqrt{2\ln 2}\left(\left(2^{\phi}+1\right)/\left(2^{\phi}-1\right)\right)$, and ϕ is the bandwidth in octaves. By selecting different center frequencies and orientations, we can obtain a family of Gabor kernels from (2.19), which can be used to extract features from an image. Given a gray-level image $f(u,v)$, the convolution of $f(u, v)$ and $\psi_{\omega,\theta}(u,v)$ is given as follows:

$$Y_{\omega,\theta}(u,v)=f(u,v)*\psi_{\omega,\theta}(u,v), \quad (2.21)$$

where * denotes the convolution operator. The convolution can be computed efficiently by performing the fast Fourier transform (FFT), then point-by-point

multiplications, and finally the inverse fast Fourier transform (IFFT). Concatenating the convolution outputs, we can produce a one-dimensional Gabor representation of the input image denoted as follows:

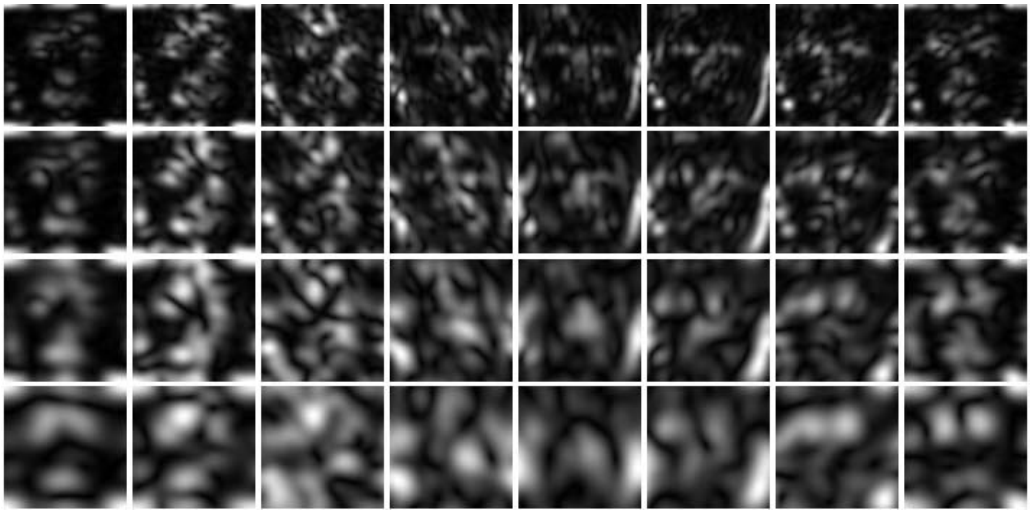
$$\mathbf{Y}_{\omega,\theta} = \left[Y_{\omega,\theta}(0,0), Y_{\omega,\theta}(0,1), \cdots, Y_{\omega,\theta}(0,N_r), Y_{\omega,\theta}(1,0), \cdots, Y_{\omega,\theta}(N_c,N_r) \right]^T, \quad (2.22)$$

where T represents the transpose operation, and N_c and N_r are the numbers of columns and rows in an image. In this thesis, we consider only the magnitude of the output of Gabor representations, which can provide a measure of the local properties of an image [54] and is less sensitive to the lighting conditions [131] (for convenience, we also denote it as $\mathbf{Y}_{\omega,\theta}$). $\mathbf{Y}_{\omega,\theta}$ is normalized to have zero mean and unit variance; and then the Gabor representations with different ω and θ are concatenated to form a high-dimensional vector for face recognition as follows:

$$\mathbf{Y} = \left[\mathbf{Y}_{\omega_1,\theta_1}^T, \mathbf{Y}_{\omega_1,\theta_2}^T, \cdots, \mathbf{Y}_{\omega_l,\theta_n}^T \right]^T, \quad (2.23)$$

where l and n are numbers of center frequencies and orientations used for the Gabor wavelets. Figure 2-1 shows the Gabor representations of a human face with 4 center frequencies and 8 orientations. It is clear that the outputs based on the Gabor wavelets exhibit strong characteristics of spatial locality, and scale and orientation selectivities.

(a)



(b)

Figure 2-1 Gabor wavelet representations of a human face. (a) The original face of size 64×64. (b) The magnitudes of the Gabor representations with 4 different center frequencies and 8 orientations. The frequencies are $\pi/2$, $\sqrt{2}\pi/4$, $\pi/4$ and $\sqrt{2}\pi/8$ from the top to the bottom row, respectively. The orientations are from 0 to $7\pi/8$ in steps of $\pi/8$, from the left to the right column, respectively.

2.1.3.6 Hausdorff Distances

Hausdorff distance is one of the shape comparison methods. This distance measure is more tolerant to perturbations in the location of points than the binary correlation techniques are. This is because the distances are measured in proximity rather than by exact superposition [51]. This method does not need to build a one-to-one pairing between the two point sets or edge maps, and only considers the spatial information about the original images. Given two finite point sets $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_n\}$, where m and n are the number of points in sets A and B. The Haudorff distance is defined as follows:

$$H(A, B) = \max\{h(A, B), h(B, A)\}, \tag{2.24}$$

where $$h(A,B)=\max_{a\in A}\min_{b\in B}\|a-b\|, \tag{2.25}$$

and $\|\cdot\|$ is an underlying norm on the point sets *A* and *B*. The function *h*(*A*, *B*) is called directed Hausdorff distance from point set *A* to *B*. For each point $a \in A$, its distance to the nearest neighbor in point set *B* is measured, and the maximum distance among the points in *A* to *B* is *h*(*A*, *B*). *h*(*B*, *A*) is computed similarly. The maximum of *h*(*A*, *B*) and *h*(*B*, *A*) is the Hausdorff distance *H*(*A*, *B*).

There are many different ways to define the distance measure *h*(*A*, *B*), so a number of modified Hausdorff distance measures have been proposed. Dubuisson et al. [52] introduced a modified Hausdorff distance (MHD), which uses the average distance instead of the maximum distance for the points in *A* when computing *h*(*A*, *B*). This can make the distance measure less sensitive to noise. The formulation of this *h*(*A*,*B*) is

$$h(A,B)=\frac{1}{N_a}\sum_{a\in A}\min_{b\in B}\|a-b\|, \tag{2.26}$$

where N_a is the number of points in set *A*. Takács [132] has proposed the "doubly" modified Hausforff distance (M2HD) for human face recognition, which is defined as follows:

$$h(A,B)=\frac{1}{N_a}\sum_{a\in A}\max\left(I\cdot\min_{b\in N_B^a}\|a-b\|,(1-I)\cdot P\right), \tag{2.27}$$

where N_B^a is the neighborhood of the point *a* in set *B*, *I* is an indicator, where *I* = 1 if there exists a point *b* within N_B^a and *I* = 0 otherwise, and *P* is an associated penalty. M2HD has been proved suitable for face recognition, where small, non-rigid local distortions are accounted for, while overall shape similarity is maintained.

2.1.3.7 Elastic Graph Matching

The dynamic link architecture (DLA) [54] is an effective face recognition approach that can handle slight perspective variations and non-rigid motion of human faces. This technique recognizes an object by using a sparse graph, where each of the vertices or nodes is labeled by a multi-resolution description in terms of a local power spectrum, and the edges of the graph are labeled by geometrical distance vectors. Object recognition can be formulated as an elastic graph matching, which is performed by minimizing a matching cost function. The local features at each vertice are extracted by using the Gabor wavelets to form a Gabor jet.

For face matching, the graph of a model face with $m \times n$ vertices is placed on the query face, and is then allowed to deform to match the query image. Let x_i, where $i = 1, \ldots, m \times n$, denote the i^{th} vertice of a graph arranged in the order from left to right and top to bottom. The Gabor jet at a vertice x_i is denoted as $J(x_i)$. The graph of the model face has its Gabor jet values equal to the Gabor wavelets representations of the image at the respective vertices. With an input image I, the model graph M is placed on it and is then allowed to deform in such a way that the cost function in (2.28) is a minimum.

$$C_{total}\left(\left\{x_i^I\right\}\right) := \lambda \sum_{(i,j)\in E} S_e\left(\vec{\Delta}_{ij}^I, \vec{\Delta}_{ij}^M\right) - \sum_{i\in V} S_v\left(J^I\left(x_i^I\right), J_i^M\right), \qquad (2.28)$$

where $\vec{\Delta}_{ij}$ is the Euclidean distance vector of the labeled edges between vertices x_i and x_j. $S_e\left(\vec{\Delta}_{ij}^I, \vec{\Delta}_{ij}^M\right)$ is a function which measures the difference between the edge labels of the image graph and the model graph. $S_v\left(J^I\left(x_i^I\right), J_i^M\right)$ is a function used to measure the similarity of the corresponding vertex labels between the image graph

and model graph, where $J^I\left(x_i^I\right)$ and J_i^M represent the i^{th} jet of the image graph and model graph, respectively. The coefficient λ is used to control the rigidity of the image graph; large value will penalize distortion of the graph *I* with respect to the graph *M*. Therefore, elastic graph matching of a model graph to an image graph in the image domain amounts to a search for a set of vertex positions, $\left\{x_i^I\right\}$ where $i = 1, \ldots, m \times n$, which optimizes the matching of the vertex labels and the edge labels.

2.2 Review of Face Recognition under Varying Illumination

2.2.1 Problem Statement

As we reviewed in Section 2.1, human face recognition, as one of the most successful applications of image analysis and understanding, has received significant attention in the last decade. However, due to difficulty in controlling the lighting conditions in practical applications, variable illumination is one of the most challenging problems with face recognition. As stated by Adini *et al*. [67], "The variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity", most existing methods for face recognition, such as PCA, and ICA, encounter difficulties under varying lighting conditions. Figure 2-2 shows some images under varying illuminations. We can see that although these images are from the same person, due to the effect of uneven lighting, they look quite different. Therefore, when the images are under varying illumination, we should build some special methods to perform the face recognition.

Figure 2-2 Samples of cropped faces from the YaleB database [133]. The azimuth angles of the lighting of images from left to right column are: 0°, 0°, 20°, 35°, 70°, -50° and -70°, respectively. The corresponding elevation angles are: 20°, 90°, -40°, 65°, -35°, -40° and 45°, respectively.

2.2.2 Literature Review

Many methods have been proposed to handle the illumination problem. The linear subspace method [42, 134-136] considered a human face image as a Lambertian surface, which can use three or more images of an object under different lighting conditions to compute a basis for the 3D illumination subspace. Without ignoring the shadows, the 3D illumination subspace model was extended to a more elaborate one, namely the illumination convex cone [66, 137, 138]. Ishiyama *et al*. [139] proposed a geodesic illumination basis model, which calculates pose-independent illumination bases for a 3D model. Batur *et al*. [140] presented a segmented linear subspace model by segmenting the images into regions that have surface normals with directions close to each other. Kouzani *et al*. [141] used an embossing technique to process a face image before presenting it to a standard face recognition system. Zhao and Yang [142] attempted to account for the arbitrary effects of illumination on PCA-based vision systems by first generating an analytically closed-form formula of the covariance matrix of faces under a particular lighting condition, and then converting it to an arbitrary illumination via an illumination equation. All the above-mentioned methods usually require a set of

known face images under different lighting conditions for training.

Zhao and Chellappa [143] developed a shape-based face recognition system by means of an illumination-independent ratio image derived by applying a symmetric shape-from-shading technique to face images. Chen *et al*. [144] adopted a probabilistic approach in which a probability distribution for the image gradient is analytically determined. Shashua *et al*. [145, 146] used quotient images to solve the problem of class-based recognition and image synthesis under varying illumination. Zhao *et al*. [32] proposed illumination ratio images, which can be used to generate new training images for face recognition with a single frontal view image. Xie *et al*. [68] proposed a model-based illumination compensation scheme for face recognition, which adopts a 2D face shape model to eliminate the effect of difference in the face shape of different persons. Liu *et al*. [147] also proposed a method that can restore a face image captured under an arbitrary lighting condition to the one with frontal illumination by using a ratio image.

2.3 Review of Facial Expression Recognition

2.3.1 Problem Statement

Over the last decade, the research on automatic facial expression analysis has become active; this has potential applications in areas such as human-computer interfaces, lip reading, face-image compression, synthetic face animation, video conferencing, human emotion analysis [20, 21], etc. Facial expressions are generated by the contractions of facial muscles, which result in the deformation of facial features such as the eyelids, eyebrows, nose and lips, and also result in changes to their relative positions. Similar muscle movements or facial feature

deformations of different identities can be arranged in a same expression model, and this process is called facial expression recognition.

2.3.2 Literature Review

The facial action coding system (FACS) [25] provides the most widely used method to measure facial movement. In the FACS, a face is divided into 44 action units (AUs) according to their locations as well as their intensities. A combination of the AUs is used to model the respective expressions. Similar coding schemes [148, 149] have also been proposed. The MPEG-4-SNHC [150] is a standard that consists of analysis, coding [151] and animation of faces (talking heads) [152]. Donato *et al*. [23] compared different techniques for the automatic recognition of facial actions, and the best performance was achieved using the Gabor wavelet representation and the independent component representation, both of which can achieve an accuracy of 96% for classifying 12 facial actions of the upper and lower face. Tian *et al*. [24] developed an automatic face analysis (AFA) system to analyze facial expressions based on both permanent facial features (brows, eyes, mouth) and transient facial features (deepening of facial furrows) in a nearly frontal-view face image sequence. This system can achieve average recognition rates of 96.4% for the upper face AUs and 96.7% for the lower face AUs. Pantic *et al*. [153] presented an automatic facial gesture recognition system based on static, frontal- and/or profile-view color face images, and a recognition rate of 86% was achieved.

Similar to FACS, a facial expression also represents the shape or position variations of the facial features between a query image and its corresponding image under normal expression. Therefore, most methods of facial expression recognition

are based on a sequence of images or a video shot, which includes face images with various expressions and images under normal expression as a reference. Donohue *et al*. [154] used the back-propagation algorithm to train a neural network, and a recognition rate of 85% based on 20 test cases was reported. Choi *et al*. [155] analyzed an input image sequence and estimated a 3D facial model, which was then used for synthesizing various facial expressions. Yacoob *et al*. [156] utilized the optical flow computation to identify the direction of rigid and nonrigid motions caused by human facial expressions, and also developed a mid-level symbolic representation motivated by psychological considerations. Huang *et al*. [157] applied a point distribution model and a gray-level model to locate the facial features, which are described by 10 action parameters (APs). For facial expression recognition, the 10 APs of a query image sequence are extracted and analyzed using PCA. Essa *et al*. [22] described a computer vision system for observing facial motion by using an optimal estimation method for optical flow, coupled with geometric, physical and motion-based dynamic models to describe a facial structure. The expression recognition accuracy was reported as 98% on a database of 52 sequences, using either the proposed muscle models or 2D motion energy models for classification. Oliver *et al*. [158] proposed a method based on 2D blob features, which are spatially compact clusters of pixels similar in terms of low-level image properties, and the HMM was adopted for facial expression and head movement classification. In [159], the HMM method was also used for recognition, while the moment invariants were used as features; the recognition rate was reported as 96.77%.

Due to the absence of reference images with a normal expression, it is more difficult to analyze facial feature actions based on a single still image, as well as to recognize the corresponding facial expression. A psychological study [160] shows that a moving display of expressions can be recognized more accurately than static images. However, for many multimedia and man-machine interface applications, such as multimedia data retrieval over the Internet, expression-based face recognition and interactive Internet games, only static images are available [161]. Therefore, in recent years, more and more attention has been focused on this field. Cottrell et al. [162] and Padgett et al. [163] used PCA to recognize facial expressions. Lyons *et al*. [164] proposed a method for classifying facial images automatically based on the labeled elastic graph matching, 2D Gabor wavelet representation, and LDA. For recognizing facial expressions, the recognition rate is 92%. Gao et al. [161] used structural and geometrical features of a user-sketched expression model to match the line edge map (LEM) descriptor of an input face image. Chen *et al*. [165] described a new feature extraction method, called clustering-based discriminant analysis (CDA), for facial expression recognition, which outperforms the traditional PCA and LDA methods. Matsugu *et al*. [166] described a rule-based algorithm combined with robust face detection using a convolutional neural network. The result shows the reliable detection of smiles, with a recognition rate of 97.6% for 5600 still images of more than 10 subjects. Abboud *et al*. [167] used an active appearance model for facial expression recognition and synthesis, which can normalize the facial expression on a given face and artificially synthesize novel expressions on the same face. Ma *et al*. [168] employed the 2D

discrete cosine transform on face images as a feature detector, and a constructive one-hidden-layer feed-forward neural network as a facial expression classifier. The best recognition rates are 100% and 93.75% (without rejection) for the training and testing images, respectively.

2.4 Conclusions

This chapter has described the principles of face recognition and facial expression recognition. We have reviewed some well-known face recognition techniques, such as PCA, LDA, ICA, Kernel PCA, Hausdorff distance, Gabor wavelets and elastic graph matching (EGM). We have also reviewed the recent development of the methods for face recognition under varying illuminations and the methods for facial expression recognition. In the chapters that follow, we will present our proposed algorithms, and compare them to those existing methods described in this chapter.

Chapter 3. Elastic Shape-Texture Matching for Human Face Recognition

In this chapter, we will present a novel elastic shape-texture matching method, namely ESTM, for human face recognition under various conditions. This method considers not only the shape information of an input image, but also adopts the corresponding texture features.

3.1 Introduction

The morphable face model [169-172] has achieved great success in encoding and representing human face images. This approach separates a given image into its shape and texture information. The shape encodes the feature geometry of the face, which is represented by a set of facial feature points and can be used to construct a pixel-wise correspondence on a reference image. The texture, which is shape-free, can be obtained after mapping the original image onto the reference image. Therefore, the shape-free texture information can be constructed only after the shape information about a face has been obtained. In other words, the first step of this approach is to detect and locate the important facial feature points. Then these points are used as control points to build a correspondence to the reference model in order to construct the shape-free texture information about the face image. Many different methods have been proposed to locate facial features [173, 104] and detect face contours [59, 174]. Although the morphable face approach has been reported

for some special applications, it is still difficult to accomplish this automatically and to achieve robust performance for images under various conditions [171].

Psychological studies have indicated that line drawings of objects can be recognized as quickly and almost as accurately as photographs [175, 176], which means that the edge-like retinal images of faces can be used for face recognition at the level of early vision. Therefore, the edges of a face image can be considered the aggregate of important feature points that are useful for face recognition. Hausdorff distance [51, 52] is such an approach, whereby the distance between two edge maps or point sets can be calculated without the explicit pairing of the points. This means that we can use Hausdorff distance to compute the similarity and perform face recognition between two edge maps. The smaller the Hausdorff distance, the smaller the difference or deformation between the two corresponding edge maps is, and the more similar the two corresponding face images are. Takács [132] has introduced a modified Hausdorff distance, which provides a more reliable and robust distance measure between two point sets than the original one. A spatially weighted modified Hausdorff distance [177] has also been proposed, which considers the importance of facial features and allocates different weights to the points according to the importance of the facial regions. Lin et al. [56] incorporates the *a priori* structure of a human face, namely eigen-mask, to emphasize the importance of facial regions and achieves a better performance level. All these methods are based on edge maps without considering any texture information about the input images.

The Gabor wavelets, whose kernels are similar to the response of the two-dimensional receptive field profiles of the mammalian simple cortical cell [53],

exhibit the desirable characteristics of capturing salient visual properties such as spatial localization, orientation selectivity, and spatial frequency [44]. The Gabor wavelets can effectively abstract local and discriminating features, which are useful for texture detection [127] and face recognition [54, 128, 129]. In [54], the Gabor wavelets have been applied for face recognition via the dynamic link architecture (DLA) framework. The DLA first computes the Gabor jets of the face images, then elastic graph matching (EGM) is used to compare their resulting image decompositions. Duc et al. [94] has introduced an automatic weighting for the nodes of the elastic graph according to their significance, and also explored the significance of the elastic deformation for an application of face-based person authentication. Kotropoulos et al. [95] has proposed a morphological dynamic link architecture which adopts discriminatory power coefficients to weigh the matching error at each grid node. Although these methods can preserve some texture features and local geometry information [96], the graph structure cannot sufficiently and effectively represent the distribution of all the feature points of human faces.

The shape and texture of a face image are complementary and supplementary to each other. Therefore, in this chapter, we propose a novel elastic shape-texture matching (ESTM) method for face recognition. Our method considers the edge map, which represents the shape information about a face image, and the Gabor wavelets, which characterize the corresponding texture information. The angles (gradient direction) of the edge points [178], which provide additional information about the shape, are also incorporated in our algorithm. Based on the shape and texture information, an elastic matching is proposed for face recognition. Unlike the

morphable face model, our method does not need to find the pixel-wise correspondence between images, which is a very difficult task in practical applications. Our algorithm, ESTM, can combine the shape, texture and angle information effectively for face recognition. Experimental results based on different databases show that ESTM outperforms other methods that employ either the shape (edge map) or the texture (Gabor wavelets) information only under various image conditions.

This chapter is organized as follows. Section 3.2 describes our proposed ESTM method. Experimental results are given in Section 3.3, which compare the performances of our proposed algorithm to other face recognition algorithms based on the Yale database, the AR database, the ORL database [179] and the YaleB database. Finally, conclusions are drawn in Section 3.4.

3.2 Elastic Shape-Texture Matching

It has been shown that the combined shape and texture feature carries the most discriminating information for human face recognition [169]. In fact, these two features are complementary to each other, and they contain the complete information about face images. We therefore propose an efficient algorithm, which combines these two types of information for face recognition. This algorithm is called Elastic Shape-Texture Matching (ESTM). In our approach, the edge map is used to represent the shape information about a face image, instead of using some specific feature points that are very difficult to locate accurately in practice. The Gabor wavelets, which exhibit strong characteristics of spatial locality and orientation selectivity, are used to extract the texture information. As only the

magnitudes of the Gabor representations are employed, we also consider the gradient direction [178] of each edge point in representing a shape. The gradient direction in this chapter is called the angle information, which is defined as follows:

$$\theta_G(x, y) = \arctan\left(\frac{f_y(x, y)}{f_x(x, y)}\right), \tag{3.1}$$

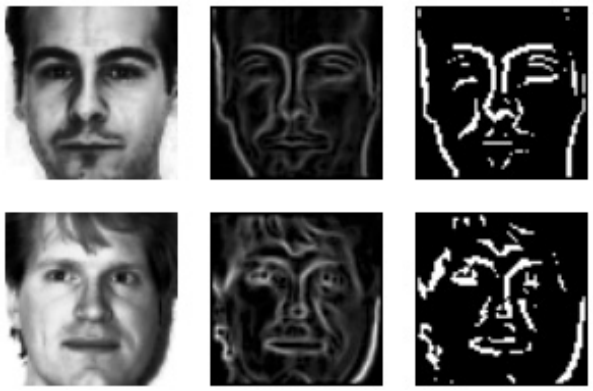
where $f_x(x, y) = f(x, y)*K_x(x, y)$, $f_y(x, y) = f(x, y)*K_y(x, y)$, $f(x, y)$ represents the gray-level intensity of an image at the coordinates (x, y), * denotes a 2D convolution operation, and $K_x(x, y)$ and $K_y(x, y)$ are the Sobel horizontal and vertical gradient kernels, respectively.

3.2.1 The Edge Maps, Gabor Maps and Angle Maps

In order to obtain the edge map of a face image, morphological operations [178] are first applied. In this thesis, the output of an image after edge detection is called an edge image, while after a thresholding procedure, the binary image produced is called an edge map of the image. The optimal parameters for an edge detector are strongly dependent on the image itself [180]; this means that a fixed threshold cannot achieve an optimal performance of converting different edge images to their corresponding binary images. In our approach, when determining the threshold to be used, we consider not only the edge image $E_G(x, y)$, but also the intensity values of the original image $f(x, y)$. This is because the important facial features, such as the eyes, mouth, etc., usually have lower gray-level intensities than other parts of a face. We define

$$n(x, y) = \frac{E_G(x, y)}{f(x, y)}. \tag{3.2}$$

Therefore, a pixel which has a larger value of $n(x, y)$ can be considered more likely to be an edge point of the facial features. The values of $n(x, y)$ are sorted in descending order, and the threshold is set so that 12% of the points with the largest magnitudes of $n(x, y)$ are selected. This threshold is obtained based on the Yale database; therefore it can be considered a trade-off of different image variations. If a query image is under normal conditions, fewer edge points are enough, while in cases of large variations, i.e. uneven lighting, more edge points are required to provide a reliable edge map. The binary edge map obtained is denoted as $E(x,y)$. Figure 3-1(b) shows the edge images obtained by the morphological edge detection, and Figure 3-1(c) displays the corresponding edge maps by using this adaptive thresholding scheme.



(a) (b) (c)

Figure 3-1 (a) The original facial images. (b) The edge images obtained by morphological operations. (c) The edge maps obtained by the adaptive thresholding method.

The Gabor map of an image is denoted as $\widetilde{G}(x, y)$, which is obtained by concatenating the Gabor wavelet representations, as shown in Figure 2-1(b), at different center frequencies and orientations. The dimension of $\widetilde{G}(x, y)$ is therefore

determined by the numbers of center frequencies and orientations used. To reduce the dimension of this representation, only one center frequency and eight orientations are considered in our algorithm. The center frequency is chosen to be $\pi/2$, and the orientation varies from 0 to $7\pi/8$ in steps of $\pi/8$.

Using (3.1), the gradient direction or angle of each point of an image can be computed. The gradient direction of a pixel varies from $-\pi/2$ to $\pi/2$. This angle information is also useful for describing the shape, and the angle map of an image is denoted as $A(x, y)$.

3.2.2 Shape-Texture Hausdorff Distance

For the edge map $E(x, y)$, Gabor map $\tilde{G}(x, y)$, and angle map $A(x, y)$, our shape-texture Hausdorff distance is defined as follows:

Given two human face images A and B, two finite point sets $A_P = \{a_1, \ldots, a_{N_A}\}$ and $B_P = \{b_1, \ldots, b_{N_B}\}$ can be obtained, where the elements in A_P and B_P correspond to the points in the edge maps E_A and E_B of the original images, and N_A and N_B are the corresponding numbers of points in sets A_P and B_P, respectively. Then, the shape-texture Hausdorff distance is

$$H(A,B) = \max(h_{st}(A,B), h_{st}(B,A)). \tag{3.3}$$

$h_{st}(A,B)$ is called the directed shape-texture Hausdorff distance, and is defined as follows:

$$h_{st}(A,B) = \frac{1}{N_A} \sum_{a \in A_P} \max\left(I \cdot \min_{b \in N_{B_P}^a} d(a,b), (1-I) \cdot P \right), \tag{3.4}$$

where $N_{B_P}^a$ is the neighborhood of the point a in the set B_P, P is an associated penalty, and I is an indicator, which is equal to 1 if there exists a point $b \in N_{B_P}^a$, and

equal to 0 otherwise. $d(a, b)$ is a distance measure between the point pair (a, b), which consists of three different terms as follows:

$$d(a, b)= \alpha \cdot d_e(a, b)+ \beta \cdot d_g(a, b)+ \gamma \cdot d_a(a, b), \tag{3.5}$$

where $d_e(a, b)$, $d_g(a, b)$ and $d_a(a, b)$ are the edge distance, Gabor distance and angle distance, respectively, for the pixel $a \in A_P$ to a pixel b within the neighborhood of a in B_P, and α, β, and γ are the coefficients used to adjust the weights of these three distance measures. All these three measures are independent of each other and are defined as follows:

$$d_e(a,b) = \|a - b\|, \tag{3.6}$$

$$d_g(a,b) = \left\|\tilde{G}_A(a) - \tilde{G}_B(b)\right\|, \text{ and} \tag{3.7}$$

$$d_a(a,b) = \|A_A(a) - A_B(b)\|, \tag{3.8}$$

where $\|.\|$ is an underlying norm, $\tilde{G}_A$, $\tilde{G}_B$, A_A, and A_B are the Gabor maps and angle maps of the two images, respectively.

In fact, the penalty P in (3.4) can also be considered as a combination of three parts, similar to (3.5), i.e.

$$P= \alpha \cdot P_e+ \beta \cdot P_g+ \gamma \cdot P_a, \tag{3.9}$$

where P_e, P_g, and P_a are the corresponding penalties for these three distance measures, respectively, and α, β, and γ have the same values as in (3.5). An advantage of using (3.9) instead of a fixed P is that this allows us to adopt different penalties for different distance measures. For example, when the lighting conditions vary significantly, some edges cannot be detected in the edge map. Due to the fact that the representations by Gabor wavelets magnitudes are less sensitive to the lighting conditions [131], we define

$$P_g(a) = \left\| \tilde{G}_A(a) - \tilde{G}_B(a) \right\|. \tag{3.10}$$

Therefore, if a point of B_P cannot be found in $N_{B_P}^a$ for the point $a \in A_P$, the corresponding Gabor representations for image B at position a will be considered when computing the penalty for Gabor distance. In other words, the value of the penalty $P_g(a)$ is adaptive to the point under consideration. This is useful to alleviate the effect of being unable to detect the edges under poor lighting. A similar mechanism can also be considered for computing P_a in some cases. As described in [144], the probability of the angles between two image gradients can serve as a measure for face recognition under varying illumination, where an empirically collected database is used to obtain the probability function. However, for a practical face recognition approach, we should consider images not only under varying illumination, but also under other conditions, such as facial expression variation and perspective variation. It is therefore difficult to obtain a proper probability function for all these cases; so we simply use a fixed value for $P_a(a)$ to compute the penalty P in our algorithm. As the penalty P is dependent on the pixel location a concerned, we use $P(a)$ instead of a fixed value P in (3.4), i.e.

$$h_{st}(A,B) = \frac{1}{N_A} \sum_{a \in A_P} \max\left(I \cdot \min_{b \in N_{B_P}^a} d(a,b), (1-I) \cdot P(a) \right). \tag{3.11}$$

3.2.3 ESTM for Face Recognition

Using the shape-texture Hausdorff distance, an elastic shape-texture matching for face recognition is proposed. For two similar face images A and B, each point in the edge point set A_P should have a corresponding near point from the edge point set B_P, with a similar texture, and vice versa. All matching pairs should fall within a

given neighborhood. In other words, this matching is non-rigid, i.e. elastic, which can tolerate small local distortions of a human face. Only edge points are considered when computing the distance, which can greatly reduce the computational complexity and memory requirement. Furthermore, the Gabor map and angle map can provide complementary discriminating information for face recognition. Therefore, this ESTM approach can be considered as a combination of template matching and geometrical feature matching [104].

As shown in (3.5), the values of $\{\alpha, \beta, \gamma\}$ are the weights of the three distance measures, which affect the recognition results. If $\alpha \neq 0$, $\beta = 0$ and $\gamma = 0$, ESTM is equivalent to M2HD. Table 3-1 shows some combinations of $\{\alpha, \beta, \gamma\}$, which will be tested in Section 3.3. For each of the combinations, the corresponding optimal set of parameters is also tabulated, where the Yale database is used as the training data.

Table 3-1 The optimal sets of parameter for different conditions of $\{\alpha, \beta, \gamma\}$.

Abbreviation	Conditions	Parameter Set
M2HD	$\alpha \neq 0, \beta = 0, \gamma = 0$	$\alpha = 1, P_e = 4.8$
ESTM$_a$	$\alpha = 0, \beta = 0, \gamma \neq 0$	$\gamma = 1, P_a = \pi/20$
ESTM$_g$	$\alpha = 0, \beta \neq 0, \gamma = 0$	$\beta = 1$
ESTM$_{ea}$	$\alpha \neq 0, \beta = 0, \gamma \neq 0$	$\alpha = 0.04, \gamma = 0.96, P_e = 4.8, P_a = \pi/20$
ESTM$_{eg}$	$\alpha \neq 0, \beta \neq 0, \gamma = 0$	$\alpha = 0.32, \beta = 0.68, P_e = 4.8$
ESTM	$\alpha \neq 0, \beta \neq 0, \gamma \neq 0$	$\alpha = 0.02, \beta = 0.05, \gamma = 0.93, P_e = 4.8, P_a = \pi/30$

3.3 Experimental Results

In this section, we will evaluate the performances of the ESTM algorithm with different conditions of the parameter set $\{\alpha, \beta, \gamma\}$ for face recognition based on different face databases. The databases used include the Yale database [193], the

AR database [194], the ORL database [179] and the YaleB database [133]. The number of distinct subjects and the number of testing images in the respective databases are tabulated in Table 3-2.

Table 3-2 The test databases used in the experiments.

	Yale	AR	ORL	YaleB
Number of subjects	15	121	40	10
Number of test images	150	605	360	640

The face images in different databases are captured under different conditions, such as varied lighting conditions, facial expressions, etc. Figure 3-2 shows some examples of the images. In order to investigate the effect of the different conditions on the face recognition algorithms, the face images in the databases are divided manually into several sub-classes according to their different conditions, and the corresponding numbers are tabulated in Table 3-3. A normal image means that the face image is of frontal view, and under even illumination and neutral expression. In our experiments, a face is under even illumination if the azimuth angle and the elevation angle of the lighting are both less than 20°. In Table 3-3, we have also combined the respective sub-classes of the same conditions to form the combined databases. For each of the combined databases, the training set consists of images from the corresponding sub-classes, e.g. the training and testing images of the combined database under normal conditions come from the Yale database, ORL database and YaleB database only.


(a)

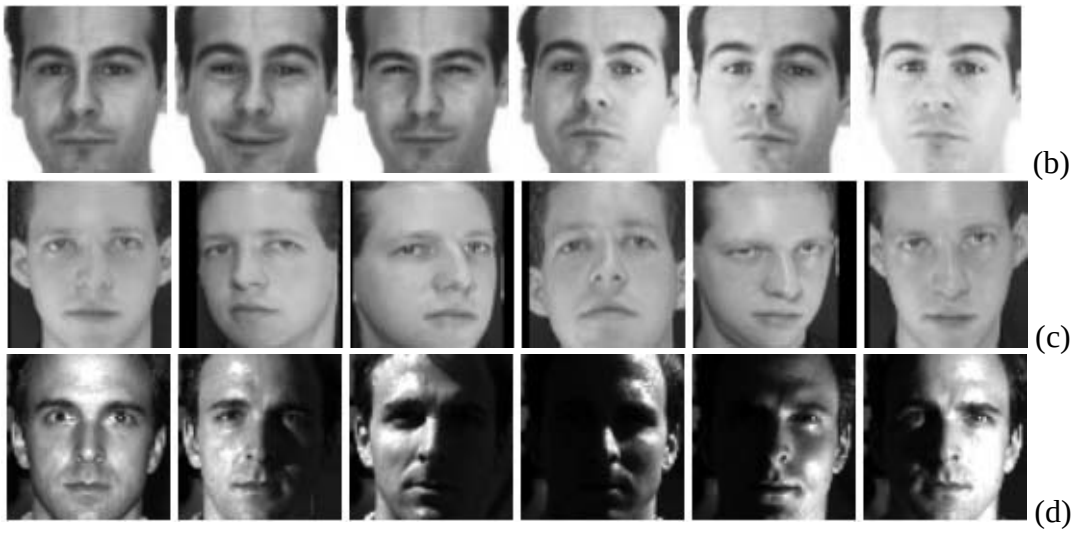


Figure 3-2 Some cropped faces used in our experiments. (a) Images from the Yale database. (b) Images from the AR database. (c) Images from the ORL database. (4) Images from the YaleB database.

Table 3-3 The sub-classes of the test databases used in the experiments.

	Normal	Facial Expression Variation	Lighting Variation	Perspective Variation
Yale	45	75	30	-
AR	-	242	363	-
ORL	189	63	-	108
YaleB	160	-	480	-
Combined	394	380	873	108

In each database, one frontal image for each subject with normal illumination and neutral expression was selected as a training sample, and others form the testing set. The respective eye locations of each image are detected and used for normalization and alignment. All images are cropped to a size of 64×64 based on the eye locations. In our system, the position of the two eyes can be located either manually or automatically [173, 181], and the input color images are converted to gray-scale ones. In order to enhance the global contrast of the images, and reduce

the effect of uneven illuminations, histogram equalization is applied to all the images. In all our experiments, we set the neighborhood size at 9×9, which is suitable for small, non-rigid local distortions in human face recognition.

The performances of our proposed ESTM and its several simplified versions, as listed in Table 3-1, are evaluated and compared with the PCA, M2HD [132], Gabor wavelets (GW), and EGM [54]. For PCA, all the eigenfaces available for each database are used, i.e. at most $M-1$, where M is the total number of training samples. In other words, 100% of the variance is kept. For example, for the Yale database, AR database, ORL database and YaleB database, 14, 120, 39 and 9 eigenfaces, respectively, are employed, respectively. For the combined databases under normal conditions, facial expression variation, lighting variation and perspective variation, the corresponding numbers of Eigenfaces used are 64, 175, 145 and 39, respectively. The GW adopts one center frequency and eight orientations, which are the same as the ESTM. For GW, the Gabor wavelets representations are concatenated to form a high-dimensional vector, which is used directly to compute the distance between two images pixel by pixel. The number of center frequencies and orientations used in EGM are five and eight, respectively, and the dimension of the elastic graph is 6×8.

3.3.1 Face Recognition Under Normal Conditions

The respective recognition rates based on the different sub-databases with normal faces are shown in Table 3-4. From the result, we can observe that:

1. Under normal conditions, most of the algorithms can also achieve a high recognition rate. In particular, PCA, GW, and ESTM can achieve a

recognition rate of 100% for the YaleB database, which contains 10 distinct subjects only. The performances of the algorithms are the worst with the ORL database, because the faces in the ORL database have some small facial expression and perspective variations.

2. The GW always outperforms the PCA, EGM and M2HD. This is consistent with the results in [169], i.e. the texture carries more discriminating information than the shape. The M2HD considers only the shape information, while the GW uses the texture information only in the matching. Although the $ESTM_g$ uses only 12% of the pixels in an image as edge points, this method can still achieve similar recognition rates to the GW with the same numbers of center frequencies and orientations for the Gabor filters. This observation shows that the edge points can be considered as the aggregate of important feature points that carry the most discriminating information for face recognition.
3. The recognition rates using $ESTM_a$ are similar to the results using M2HD. Furthermore, $ESTM_{ea}$, which adopts not only the edge information, but also the angle information, can achieve a better performance than both $ESTM_a$ and M2HD in most cases. Therefore, the angle information can be considered a complementary feature to the edge map.
4. Our proposed ESTM method, which combines the edge information, texture information and angle information, always outperforms other methods. This shows that the combined features carry the most discriminating information, rather than using only one or two of them.

Table 3-4 Face recognition results under normal conditions.

Recognition Rate (%)	PCA	GW	EGM	M2HD	$ESTM_a$	$ESTM_g$	$ESTM_{ea}$	$ESTM_{eg}$	ESTM
Yale	82.2	88.9	73.3	80.0	91.1	86.7	93.3	88.9	93.3
ORL	64.0	82.0	72.5	79.4	77.8	84.1	79.9	84.7	84.7
YaleB	100.0	100.0	98.1	99.4	98.1	99.4	99.4	99.4	100.0
Combined	80.2	89.8	81.0	86.8	87.3	90.6	88.8	91.1	91.4

3.3.2 Face Recognition Under Varying Lighting Conditions

The experimental results based on the images under varying lighting are shown in Table 3-5. In the Yale database, the lighting is either from the left or the right of the face images. In the AR database, besides the lighting from the left and the right, lighting from both sides of a face is also adopted. The YaleB database, which consists of 10 people and each person has 65 images with different lighting conditions, is often used to investigate the effect of lighting on face recognition. In this part of the experiments, we select only those images with obviously uneven illuminations as the testing images. In other words, only those images with azimuth angles or elevation angles of lighting larger than 20° are considered. Consequently, for each subject in the YaleB database, only 48 different illumination models are chosen for testing.

The performance of PCA degrades significantly compared to the results based on normal faces. The recognition rate based on the combined database falls from 80.2% to 45.1%. The major idea of PCA is to represent faces with their principal components. However, the variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations

due to change in face identity [67]. Hence, PCA cannot represent a face under severe lighting variations or perspective variations.

The edge map can serve as a robust representation to illumination changes if the objects concerned have sharp edges only. However, for objects with smooth surfaces, such as human faces, some of the edges may not be detected in a consistent manner [182]. Moreover, when the lighting is not from the front of a face, the shadows produced will also affect the edge map generated. Therefore, in the case of large illumination variation, such as the YaleB database, the performances of those algorithms that rely on the edge information for recognition, such as the M2HD, will be greatly affected. When the lighting conditions are not so poor, e.g. in the Yale database or the AR database, ESTM$_a$ always outperforms M2HD. This result is consistent with the conclusion in [144] that the direction of the image gradient is insensitive to changes in illumination direction.

It is interesting to note that the GW can still obtain a very high recognition rate, which is 97.9%, based on the YaleB database. This shows that the Gabor wavelets representations can effectively reduce the effect of varying illuminations. The EGM also adopts the Gabor representations. However, this approach uses a limited number of Gabor jets and a deformed graph in its representation. When the lighting is under very poor conditions, its recognition performance becomes poor, even poorer than that of the PCA and M2HD.

ESTM outperforms other algorithms in most cases, except when the YaleB database is used. In this case, the GW performs the best. This is due to the fact that ESTM also employs the edge map; this representation becomes inaccurate under

poor lighting. However, for the Yale database and the AR database, where the illumination variations are not large, the ESTM outperforms the GW. Furthermore, compared to the results of the M2HD which is also based on edge maps, the ESTM can achieve higher recognition rates of 13.2% to 26.8%.

Table 3-5 Face recognition results under varying lighting conditions.

Recognition Rate (%)	PCA	GW	EGM	M2HD	ESTM$_a$	ESTM$_g$	ESTM$_{ea}$	ESTM$_{eg}$	ESTM
Yale	46.7	73.3	83.3	76.7	90.0	76.7	90.0	83.3	90.0
AR	80.4	94.5	71.3	84.0	94.2	93.7	96.4	94.8	97.2
YaleB	60.8	97.9	50.0	59.2	57.1	77.7	68.0	81.5	86.0
Combined	45.1	74.7	42.3	49.8	55.6	59.7	62.1	63.0	65.5

3.3.3 Face Recognition Under Different Facial Expressions and Perspective Variations

Experiments based on the face images under different facial expressions are performed and the recognition results are summarized in Table 3-6. The performance of the GW degrades as compared to the results in Section 3.3.1. Furthermore, its recognition rate is lower than others in some cases. This is because facial expressions often cause some local distortions of the feature points, which will then affect the corresponding local texture and shape properties. The GW considers the texture information about the neighborhood of each pixel, which is disturbed by local distortions caused by changes in facial expression.

The PCA uses the principal components to represent the face images, which are less sensitive to small local distortions. Therefore, the performance of PCA also degrades in this case, but to a less extent when compared to the GW. Both the EGM and M2HD adopt elastic matching techniques, which search for matching pairs

within a neighborhood, and can therefore partially alleviate the effect of local distortions caused by facial expressions.

As compared to other methods, our proposed ESTM can achieve the best performance. The elastic matching adopted in ESTM can effectively reduce the effect of small and non-rigid local distortions caused by changes in facial expression. Moreover, the Gabor features and angle information can provide complementary information for face recognition. The recognition rate of $ESTM_g$ is slightly higher than that of the ESTM when using the AR database, and both methods have a recognition rate higher than 98%.

Table 3-6 Face recognition results under different facial expressions.

Recognition Rate (%)	PCA	GW	EGM	M2HD	$ESTM_a$	$ESTM_g$	$ESTM_{ea}$	$ESTM_{eg}$	ESTM
Yale	66.7	73.3	57.3	66.7	77.3	78.7	78.7	84.0	85.3
AR	84.3	92.1	92.1	89.7	97.5	98.8	97.1	97.5	98.3
ORL	71.4	66.7	69.8	84.1	77.8	76.2	79.4	88.9	90.5
Combined	76.3	79.7	74.2	78.2	86.6	86.8	86.8	89.2	90.0

The relative performances of the different algorithms were also evaluated for faces under perspective variations. All the testing images are selected from the ORL database with the faces either rotated out of the image plane, e.g. looking to the right, left, up and down, or rotated in the image plane, clockwise or anti-clockwise. The experimental results are tabulated in Table 3-7, and show that none of these face recognition methods can achieve a satisfactory performance under perspective variations. Nevertheless, the ESTM still outperforms the other methods.

Table 3-7 Face recognition results under various perspectives.

Recognition Rate (%)	PCA	GW	EGM	M2HD	$ESTM_a$	$ESTM_g$	$ESTM_{ea}$	$ESTM_{eg}$	ESTM
ORL	39.8	56.5	42.6	43.5	50.9	56.5	48.1	57.4	60.0

3.3.4 Face Recognition with Different Databases

We have evaluated and discussed the effect of different conditions on different face recognition methods. In this section, we also show the performances of the respective face recognition methods based on the different databases without dividing them into sub-databases. The recognition results are tabulated in Table 3-8, and also show that the ESTM outperforms all the other methods based on the different databases, except for the YaleB database. In this case, the GW achieves the best performance. In addition, the simplified versions of ESTM, i.e. $ESTM_a$, $ESTM_g$, $ESTM_{ea}$ and $ESTM_{eg}$, also outperform the traditional methods, such as PCA, GW, EGM and M2HD, in most of the cases. With these four databases, the recognition rate for the ORL database is always lower than the others, no matter which method is adopted. This is due to the effect of perspective variations, which has been discussed in Section 3.3.3.

Table 3-8 Face recognition results based on different databases.

Recognition Rate (%)	PCA	GW	EGM	M2HD	$ESTM_a$	$ESTM_g$	$ESTM_{ea}$	$ESTM_{eg}$	ESTM
Yale	67.3	78.0	67.3	72.7	84.0	80.7	85.3	85.3	88.7
AR	82.0	93.6	79.7	86.3	95.5	95.7	96.7	95.9	97.7
ORL	58.1	71.7	63.1	69.4	69.7	74.4	70.3	77.2	78.3
YaleB	70.6	98.4	62.0	69.2	67.3	83.1	75.9	85.9	89.5

3.3.5 Storage Requirements and Computational Complexity

For our approach, the data stored in a database for a face image include its edge map, Gabor map, and angle map. Suppose that the size of the normalized face is $N \times N$, and η percent of the points are selected as edge points in the edge map. The average number of feature points for an edge map is $\eta \cdot N^2$, where a feature point is the *x*- and *y*-coordinates, and can be represented by two bytes. The dimensions of the Gabor map and angle map are $n_f n_a N^2$ and N^2, respectively, where n_f and n_a are the numbers of center frequencies and orientations used for the Gabor filters. Each element in the Gabor map and the angle map is represented by a 16-bit floating-point number. Therefore, the total number of bits used to represent a face image in the database is $16\left(\eta + n_f n_a + 1\right)N^2$.

For a query image, the computational time for face recognition includes two parts: feature extraction and matching. The runtime required for feature extraction is the time spent computing the corresponding edge map, Gabor map, and angle map. As all these maps of the training images have been computed and stored in the face database, we only need to consider the time required to compute these maps of the query image. The computational complexities for computing an edge map, Gabor map and angle map are in the order of $O\left(N^2\right)$, $O\left(N^2 \log_2\left(N^2\right)\right)$ and $O\left(N^2\right)$, respectively. For searching in a large database, the runtime for matching is the most significant part for the whole process. Suppose that the size of the neighborhood considered when searching for a matching pair is $D \times D$. This means that the possible number of pixels to be compared when matching each point pair is D^2. In this

matching, the edge distance $d_e(a, b)$, Gabor distance $d_g(a, b)$ and angle distance $d_a(a, b)$ between pixel $a \in A$ and pixel $b \in B$ are to be computed. Suppose that the average runtimes required to compute these three distances for one point pair (a, b) are t_e, t_g and t_a, respectively, and that the total runtime $t_{all} = t_e + t_g + t_a$, then the computational complexity of ESTM is in the order of $O\left(2\eta N^2 D^2 M t_{all}\right)$, where a factor of 2 is multiplied, since both $h_{st}(A, B)$ and $h_{st}(B, A)$ in (3.3) are to be computed, and M is the number of images stored in the database. Experiments were conducted on a computer system with Pentium IV 2.4GHz CPU and 512MB RAM. The average runtime using ESTM for face recognition based on the ORL database (40 face subjects) is 0.6s.

3.4 Conclusions

In this chapter, we have proposed a novel elastic shape-texture matching algorithm, namely ESTM, for human face recognition. In our approach, the edge map is used to represent the shape information about an input image, and the Gabor wavelets are employed to characterize the corresponding texture information. The gradient direction can also provide additional discriminating information, which is called angle information. For a query image, its edge map, Gabor map and angle map are first computed, and then a shape-texture Hausdorff distance is proposed to compute the difference between a query input and the faces in a database. This method does not need to construct a precise pixel-wise correspondence between the two images to be compared, and the matching is performed within a neighborhood. This makes this approach robust to small and local distortions of the facial feature points, and suitable for face recognition.

This chapter also addresses the performances of different face recognition algorithms in terms of changes in facial expressions, uneven illuminations, and perspective variations. Experiments were conducted based on different databases, which show that our algorithm can always achieve the best performance as compared to other algorithms, such as PCA, GW, EGM and M2HD, under different conditions. The only exception is when the face images are under very poor lighting conditions, in which case the GW performs the best while the ESTM achieves the second highest recognition rate. With our approach, the recognition rates based on the Yale database, AR database, ORL database and YaleB database are 88.7%, 97.7%, 78.3% and 89.5%, respectively.

The ESTM method proposed in this chapter can achieve a high performance level under different conditions. However, the method requires the use of an edge map. Under severe lighting conditions, it is difficult to obtain a faithful edge map, and so the performance will degrade. In the next chapter, we will present another of our proposed face recognition algorithms, which is based on a Doubly nonlinear mapping Kernel PCA (DKPCA). DKPCA, which employs Gabor features and does not need an edge map, is more robust for lighting variations.

Chapter 4. Gabor-Based Kernel PCA with Doubly Nonlinear Mapping for Face Recognition

In Chapter 3, the method for recognizing human face images is based on the shape and texture information. This is mainly an edge-based method. When the input image is under varying illumination, an accurate edge map cannot be obtained and its performance will degrade. In this chapter, we will propose a novel Gabor-based kernel PCA with doubly nonlinear mapping method, which is robust to the image variations caused by the illumination conditions, facial expressions and perspectives.

4.1 Introduction

Although kernel-based methods [48-50, 80-84] can overcome many limitations of linear transformation, He et al. [79] pointed out that none of these methods explicitly considers the structure of the manifold on which the face images possibly reside. In this chapter, we propose a novel method for face recognition, which uses a single image per person for training, and is robust to lighting, expression and perspective variations. In our method, the Gabor wavelets are used to extract facial features, then a Doubly nonlinear mapping Kernel PCA (DKPCA) is proposed to perform the feature transformation and face recognition. Doubly nonlinear mapping means that, besides the conventional kernel function, a new mapping function is also defined and used to emphasize those features, which have

higher statistical probabilities and spatial importance for face images. More specifically, this new mapping function considers not only the statistical distribution of the Gabor features, but also the spatial information about human faces. After this nonlinear mapping, the transformed features have a higher discriminating power, and the importance of the features adapts to the spatial importance of the face images. Therefore, it has the ability to reduce the effect of feature variations due to illumination, expression and perspective disturbance. We evaluate the performance of the proposed algorithm for face recognition with the use of different databases, which in total involve 186 identities and 1755 testing images produced under various conditions. Consistent and promising results were obtained, which show that our method can greatly improve recognition performances in all conditions.

This chapter is organized as follows. Section 4.2 describes our new doubly nonlinear mapping kernel PCA. Experimental results are given in Section 4.3, which compare the performances of our proposed algorithm to other face recognition algorithms based on the Yale database, the AR database, the ORL database and the YaleB database. Finally, conclusions are drawn in Section 4.4.

4.2 Doubly Nonlinear Mapping Kernel PCA

As described in Section 2.1.3.4, although we do not need to perform the nonlinear mapping explicitly in KPCA, and all the computations are implemented in the input space instead of the high-dimensional feature space, it is still meaningful to investigate how to design a “good” mapping that has an explicit physical meaning and is suitable for pattern recognition applications, such as face recognition. In fact, as mentioned in [79], none of the kernel-based methods explicitly considers the

structure of the manifold on which the face images possibly reside. In this section, we will propose a novel KPCA with doubly nonlinear mapping, which considers not only the statistical property of the input Gabor features, but also the spatial information about human faces.

In traditional KPCA, kernel tricks are employed to compute the dot products in the original low-dimensional input space by means of a kernel function [47, 48]. From (2.16) to (2.18), three classes of kernel functions which are widely adopted, we can see that whichever kernel function is used, the input *N*-dimensional variable **Y** is holistically considered. In other words, each element $y \in \mathbf{Y}$ is treated equally and acts in the same role. However, due to the uneven statistical probability of *y* and the different spatial importance in a face image, the elements with different values and spatial locations should be assigned different weights for discrimination. In our approach, the statistical probability distribution of *y* is approximated by a normal density function, and an element with a higher probability should provide more discriminant information for recognition. In addition, the elements derived from the important facial features such as eyes, mouth, nose, etc., should also be emphasized. The spatial importance can be measured by means of the eigenmask **E** [55, 56], which is shown in Figure 4-1. Therefore, nonlinear mapping, Ψ, is devised to emphasize those features that have both higher statistical probabilities and spatial importance:

$$\begin{aligned} \Psi: \quad & R^N \to R^N, \\ & \mathbf{Y} \xrightarrow{\mathbf{E}} \Psi(\mathbf{Y}). \end{aligned} \tag{4.1}$$

Figure 4-1 The eigenmask used in our method.

This mapping is operated in the original input space, and **Y** has the same dimension as $\Psi(\mathbf{Y})$. For each $y \in \mathbf{Y}$, there is a corresponding *z* in the transformed feature space, i.e. $z = \Psi(y)$. The spatial importance of *y* is determined by the value of the eigenmask *s* at a pixel position, i.e. $s = \mathbf{E}(u, v)$, where *u*, *v* are the coordinates of a pixel, and *y* is the corresponding Gabor representation for the same pixel. The same value of *y* with a different *s* should have a different mapped value. In other words, *z* at the pixel position is determined by its Gabor representation *y* and its eigenmask value *s*. Therefore, we have

$$y \to z = \Psi(y, s). \tag{4.2}$$

As the statistical property of the Gabor representation and the spatial information about faces are complementary to each other, and y and *s* are independent of each other, Ψ can be represented as follows:

$$\Psi(y, s) = \Psi_1(y) \cdot \Psi_2(s). \tag{4.3}$$

The mapping is therefore the product of two nonlinear mapping functions. This has the advantage that Ψ_1 and Ψ_2 can be designed independently according to their respective properties.

For face recognition, the difference between a query input and the faces in a database is computed, and the input is assigned to the one that has the minimal difference. Therefore, the following total differential equation is considered:

$$\Delta\Psi = \frac{\partial\Psi}{\partial y}\cdot\Delta y + \frac{\partial\Psi}{\partial s}\cdot\Delta s\,. \tag{4.4}$$

With (4.3), we have

$$\begin{aligned}\Delta\Psi &= \Psi_2(s)\cdot\frac{\partial\Psi_1(y)}{\partial y}\cdot\Delta y + \Psi_1(y)\cdot\frac{\partial\Psi_2(s)}{\partial s}\cdot\Delta s \\ &= \Psi_2(s)\cdot\frac{\partial\Psi_1(y)}{\partial y}\cdot\Delta y,\end{aligned} \tag{4.5}$$

where $\Delta s = 0$ because the eigenmask E is generated based on a set of training images and is supposed to be a fixed structure.

Firstly, we consider the characteristics of Ψ_1, which maps the Gabor representations of a face image in a nonlinear manner. Each input **Y** is normalized to have zero mean and unit variance. By the central limit theorem [183], we can use a normal distribution to estimate the probability density function (PDF) $p(y)$, where y represents an element of **Y**. The nearer the value of y to the mean or zero for demeaned vectors, the more likely it is that the elements will be the expected pattern, and the more important will be their role for recognition. Therefore, the mapping function Ψ_1 should satisfy the following condition.

Condition 1: $$\frac{\partial\Psi_1(y)}{\partial y} \propto p(y)\,. \tag{4.6}$$

By (4.5) and (4.6), we have

$$\Delta\Psi \propto p(y)\cdot\Delta y\,, \tag{4.7}$$

which implies that after the nonlinear mapping Ψ_1, the feature variation Δy of the element *y* with a higher statistical probability should be given a larger weight, and so act in a more important role for discrimination, and *vice versa*.

Secondly, the nonlinear mapping Ψ_2 is based on the spatial information about human faces. The spatial information is represented by an eigenmask, which is a modification of the first eigenface derived from a set of training images [55, 56] and which is normalized between [0,1]. The higher the magnitude of an element *s* in the eigenmask, the more important the feature point it represents. Hence, Ψ_2 should satisfy the following condition.

Condition 2: Ψ_2 and its derivative function $\frac{\partial \Psi_2}{\partial s}$ are monotonically increasing functions.

In Condition 2, Ψ_2 is monotonically increasing so that those pixels belonging to the important facial features will be emphasized. Furthermore, the increase of Ψ_2 is nonlinear and at a higher rate when *s* has a higher value. Therefore, $\frac{\partial \Psi_2}{\partial s}$ should also be a monotonically increasing function.

From (4.5), we can see that

$$\Delta \Psi \propto \Psi_2(s) \cdot \Delta y, \tag{4.8}$$

which means that after the mapping Ψ_2, the feature variation Δy at an important facial point should be enhanced, and *vice versa*. In other words, the important facial features will provide more discriminant information for distinguishing two face

images. This is coincident with the fact that the facial features should be assigned different weights according to their importance for face recognition [55, 56].

From Condition 1, the differential of Ψ_1 is directly proportional to a normal distribution with zero mean and unit variance, i.e. $N(0,1)$. More generally, we have

$$\frac{\partial \Psi_1(y)}{\partial y} = N(0,a) = \frac{1}{a\sqrt{2\pi}} e^{-(y^2)/(2a^2)}, \tag{4.9}$$

where $N(0,a)$ is a normal distribution with zero mean and a variance of a, and a is a positive constant. Then, we have

$$\begin{aligned} \Psi_1(y) &= \frac{1}{a\sqrt{2\pi}} \int_{-\infty}^{y} e^{-(t^2)/(2a^2)} dt + \xi \\ &= \frac{1}{a\sqrt{2\pi}} \int_{0}^{y} e^{-(t^2)/(2a^2)} dt + 0.5 + \xi \\ &= \frac{1}{2} erf\left(\frac{y}{\sqrt{2}a}\right) + 0.5 + \xi, \end{aligned} \tag{4.10}$$

where ξ is a constant and $erf(y)$ is the "error function" which is the integration of the normal distribution [184] and is defined as follows:

$$erf(y) \equiv \frac{2}{\sqrt{\pi}} \int_{0}^{y} e^{-t^2} dt. \tag{4.11}$$

As it is desirable that the data after the mapping should also be centered, as required for performing the KPCA [48] in the next step, ξ is set at –0.5.

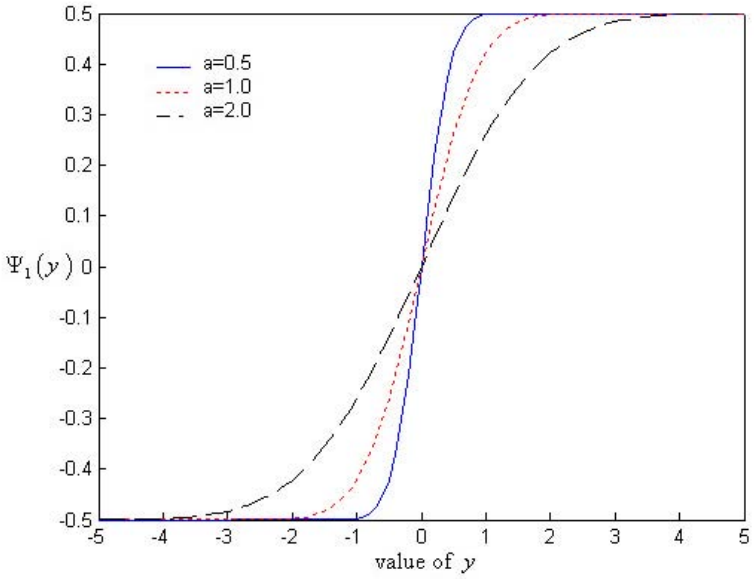
From Condition 2, Ψ_2 and $\frac{\partial \Psi_2}{\partial s}$ are monotonically increasing, therefore we set Ψ_2 as follows:

$$\Psi_2(s) = b^s, \tag{4.12}$$

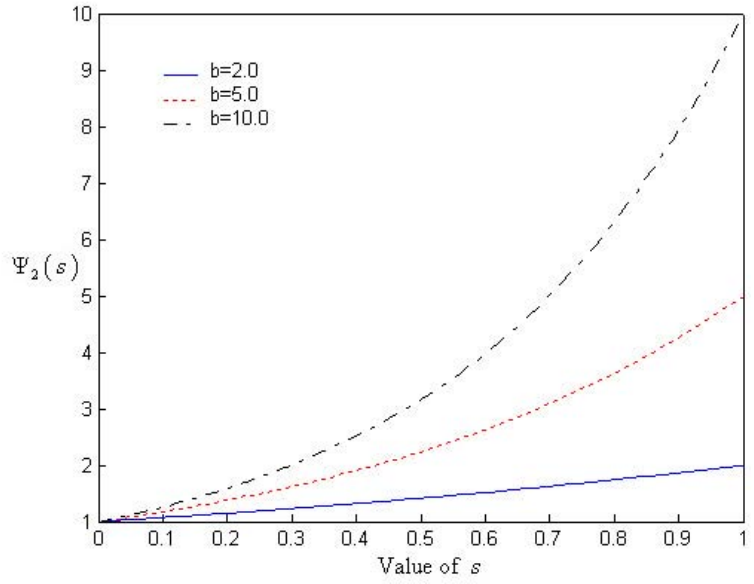
where $b > 1$. Considering (4.2), (4.3), (4.10) and (4.12), we have the doubly nonlinear mapping function as follows:

$$z = \frac{1}{2} erf\left(\frac{y}{\sqrt{2}a}\right) \cdot b^{s}. \tag{4.13}$$

Figures 4-2(a) and (b) illustrate the graphs of Ψ_1 and Ψ_2 with different values of *a* and *b*, respectively.



(a)



(b)

Figure 4-2 (a) The graph of function $\Psi_1(y)$ with different values of the parameter *a*, and (b) the graph of function $\Psi_2(s)$ with different values of the parameter *b*.

After this nonlinear mapping, an element in **Y**, which has a higher statistical probability and spatial importance, can act in a more important role for face recognition. This mapping takes place in the input space and does not increase the data dimensionality. Combined with the conventional KPCA, we can obtain a novel 'doubly' nonlinear mapping KPCA. This process is equivalent to performing two nonlinear mappings – the first nonlinear mapping is Ψ as shown in (4.1), and is then followed by the nonlinear mapping Φ as shown in (2.10) – on an input feature $\mathbf{Y}$ to a high-dimensional feature space, and then performing PCA for recognition. (Certainly, the second mapping Φ is not explicitly processed and all procedures are implemented in the original space, as discussed in Section 2.1.3.4.) Combining (2.10) and (4.1), the doubly nonlinear mapping KPCA defines a nonlinear mapping $\Phi(\Psi)$ as follows:

$$\begin{aligned} \Phi(\Psi): \quad & R^N \to R^N \to F, \\ & \mathbf{Y} \xrightarrow{\mathbf{E}} \Psi(\mathbf{Y}) \to \Phi(\Psi(\mathbf{Y})), \end{aligned} \tag{4.14}$$

and PCA is performed in the mapped feature space for recognition.

4.3 Experimental Results

In this section, we will evaluate the performances of the proposed doubly nonlinear mapping KPCA for face recognition based on different face databases. The databases used include the Yale database, the AR database, the ORL database, and the YaleB database. The number of distinct subjects and the number of testing images in the respective databases are tabulated in Table 4-1.

The face images in the different databases are captured under different conditions, such as varied lighting conditions, facial expressions and perspectives.

As shown in Table 3-3, the face images in the databases are divided manually into several sub-classes according to their different properties. In this chapter, we also adopt these sub-databases to investigate the effect of the different conditions on the face recognition algorithms.

Table 4-1 The test databases used in the experiments.

	Yale	AR	ORL	YaleB
Number of subjects	15	121	40	10
Number of test images	150	605	360	640

In each database, one frontal image of each subject with normal illumination and neutral expression is selected as a training sample, and the rest form the testing set. All images are cropped to a size of 64×64 based on the eye locations. In our system, the position of the two eyes can be located either manually or automatically [55, 173, 181], and the eye locations are then used for normalization and alignment. The input color images are converted to gray-scale ones. To enhance the global contrast of the images and reduce the effect of uneven illuminations, histogram equalization is applied to all the images.

Our method is to perform an additional nonlinear mapping for the conventional KPCA. In this chapter, we select the KPCA with fractional power polynomial (FPP) models [50], and evaluate its performance with and without use of the proposed doubly nonlinear mapping for face recognition. The polynomial kernel (2.16) is used, and the power is set at 0.8. To derive the real features of KPCA (2.15), we apply only those KPCA eigenvectors that are associated with positive eigenvalues. Furthermore, (2.16) is modified as

$$k\left(\mathbf{Y}_i, \mathbf{Y}_j\right) = sign\left(\mathbf{Y}_i \cdot \mathbf{Y}_j\right) \cdot \left(\left|\mathbf{Y}_i \cdot \mathbf{Y}_j\right|\right)^d, \tag{4.15}$$

where *sign*( ) is a signum function. As discussed in [50, 185], a PCA classifier will perform better when the Mahalanobis distance is used. Therefore, in our experiments, the Mahalanobis distance is also employed as the distance measure.

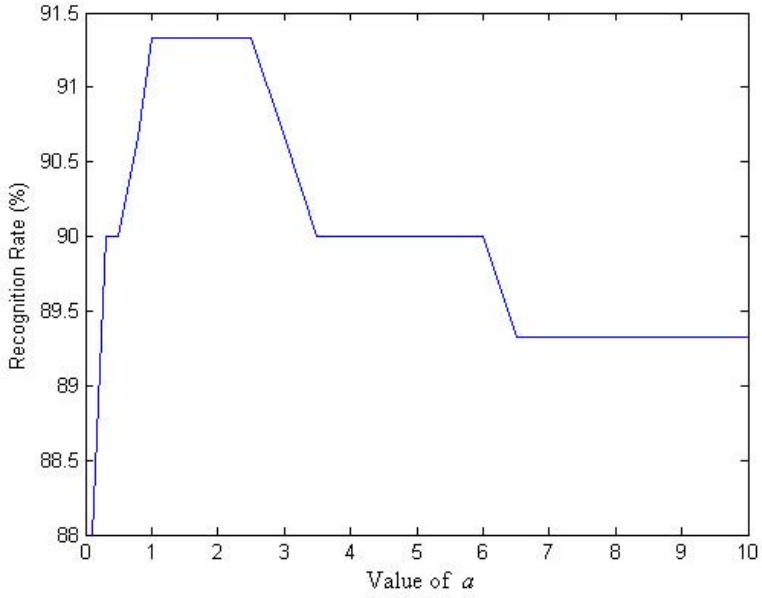
4.3.1 Determination of Parameters for the Nonlinear Mapping Functions

In (4.13), two parameters, *a* and *b*, are involved in the mapping function. From Figure 4-2, we can see that nonlinear mapping functions with different parameter values have different properties. Therefore, proper values for the parameters are to be determined so as to obtain an optimal result. In our experiments, we use the Yale database for training and determining the optimal values for *a* and *b*, and the mapping functions Ψ_1 and Ψ_2. Then, these mapping functions are evaluated using other databases. To obtain the optimal values for *a* and *b*, different values of *a* and *b* are tested, and then DKPCA is employed for face recognition. If only Ψ_1 is considered, the value of *b* in (4.13) is set at 1; and if only Ψ_2 is used, (4.13) is changed to the following form,

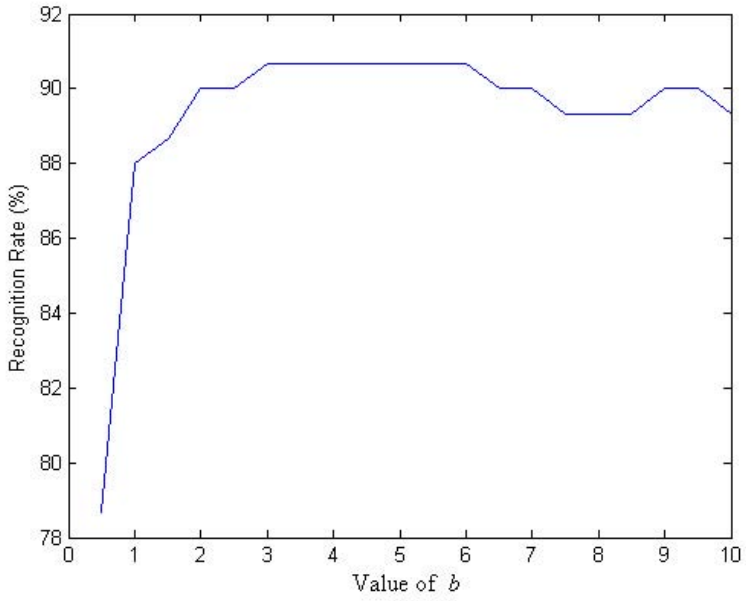
$$z = y \cdot b^s. \tag{4.16}$$

Experimental results are shown in Figure 4-3, where we can see that the best performance is achieved if the values of *a* and *b* are set at [1.0, 2.5] and [3, 6] when only Ψ_1 and only Ψ_2, respectively, are employed in face recognition. When we consider these two parameters at the same time, i.e. $\Psi = \Psi_1 \cdot \Psi_2$ is used for nonlinear feature transformation, experimental results show that they should be set at $a = 1.0$ and $b = 3.0$ for the best performance. This result coincides with the discussion in

Section 4.2. With Condition 1, when $\frac{\partial \Psi_1(y)}{\partial y} \propto p(y)$ is satisfied, the optimal transformation for face recognition can be achieved. Considering the assumption that $p(y) \propto N(0,1)$, the value of the parameter *a* should be close to 1.



(a)



(b)

Figure 4-3 Face recognition using (a) nonlinear mapping Ψ_1 with different values of *a* and (b) nonlinear mapping Ψ_2 with different values of *b*.

4.3.2 Face Recognition Under Normal Conditions

The respective performances of several face recognition methods based on the normal faces from the different databases are shown in Table 4-2. GW+PCA means using Gabor representations as the facial features, and then adopting PCA to reduce the feature dimension and perform face recognition. GW+KPCA is the KPCA with FPP proposed in [40]. GW+DKPCA$_1$, GW+DKPCA$_2$ and GW+DKPCA represent our doubly nonlinear mapping KPCA, and the mapping functions used are Ψ_1, Ψ_2 and Ψ, respectively. From the result, we can observe that:

1. Under normal conditions, most of the algorithms can achieve a high recognition rate. In particular, all the methods considered can achieve a recognition rate of 100% for the YaleB database, which contains 10 distinct subjects only. The performances of the algorithms are the worst with the ORL database, because the faces in the ORL database have some small variations in facial expression and perspective.
2. The Gabor-based methods outperform the PCA method. This is because Gabor filters can extract detailed local textures, which exhibit the desirable characteristics of capturing salient visual properties such as spatial localization, orientation selectivity, and spatial frequency, while the PCA method mainly focuses on maintaining the global structure of training images, and is not optimal for discrimination.
3. The performance of the KPCA with FPP can be the same but is sometimes worse than the conventional Gabor-based PCA. This is due to the fact that the optimal value of the power d in (2.16) is obtained based on a combined

database, which includes more than 1,000 images under various conditions. This method may not consistently perform better for the respective databases.

4. Our proposed doubly nonlinear mapping KPCA outperforms all other methods, regardless of which mapping function is used. The method using only Ψ_1 performs better than that using only Ψ_2. This is because the latter applies a fixed eigenmask to all images, while the former transforms the inputs according to their probability distribution. Therefore, the method based on the statistical property of the input is more elastic and suitable for human face recognition. When the statistical characteristic and the spatial information are considered together, i.e. Ψ is used as the mapping function, the best performance can be achieved. This implies that these two kinds of information are complementary to each other.

Table 4-2 Face recognition results under normal conditions.

Recognition Rate (%)	PCA	GW+ PCA	GW+ KPCA	GW+ $DKPCA_1$	GW+ $DKPCA_2$	GW+ DKPCA
Yale	91.1	93.3	93.3	93.3	93.3	93.3
ORL	80.4	85.2	84.7	88.4	86.8	89.4
YaleB	100.0	100.0	100.0	100.0	100.0	100.0

4.3.3 Face Recognition Under Varying Lighting Conditions

In the Yale database, the lighting is either from the left or the right of the face images. In the AR database, besides the lighting from the left and the right, lighting from both sides of each face is also adopted. The YaleB database, which consists of 10 people with each person having 65 images under different lighting conditions, is often used to investigate the effect of lighting on face recognition. In this part of the experiments, we select only those images with obviously uneven illuminations as

the testing images. In other words, only those images with azimuth angles or elevation angles of lighting larger than 20° are considered. Consequently, for each subject in the YaleB database, only 48 different illumination models are chosen for testing. The experimental results based on the images under varying lighting from different databases are shown in Table 4-3.

The performance of PCA degrades significantly compared to the results based on the normal faces. PCA represents faces with their principal components, but the variations between the images of the same face due to illumination are almost always larger than image variations due to change in face identity [67]. Hence, PCA cannot represent and discriminate a face under severely uneven lighting conditions.

Compared to the PCA method, the Gabor wavelets can greatly increase the recognition performance based on the different databases. This shows that the Gabor wavelets representations can effectively reduce the effect of varying illumination. For the Yale database and the YaleB database, KPCA with FPP outperforms the conventional Gabor-based KPCA but the latter can achieve a better performance for the AR database.

In most cases, our doubly nonlinear mapping KPCA outperforms other algorithms, except for the YaleB database, where the KPCA with FPP performs better than our method when only Ψ_2 is used as the mapping function. This is due to the fact that Ψ_2 is derived from the eigenmask, which emphasizes the important features in a human face under normal conditions. With severe illumination variations, such as those shown in Figure 3-2(d), the Gabor features abstracted from some feature points may not be reliable for face recognition. Emphasizing these

features may result in adverse performance. However, when combined with Ψ_1, the spatial information can still provide additional and useful information and improve the recognition rate. Like the results in Section 4.3.2, DKPCA with Ψ_1 outperforms DKPCA with Ψ_2, and DKPCA with Ψ achieves the best performance.

Table 4-3 Face recognition results under varying lighting conditions.

Recognition Rate (%)	PCA	GW+ PCA	GW+ KPCA	GW+ $DKPCA_1$	GW+ $DKPCA_2$	GW+ DKPCA
Yale	60.0	90.0	93.3	96.7	93.3	100.0
AR	81.3	96.4	96.1	98.9	97.3	98.9
YaleB	53.3	91.7	94.2	97.9	92.7	98.1

4.3.4 Face Recognition with Variations in Facial Expressions and Perspective

Experiments based on the face images with different facial expressions are conducted and the recognition results are summarized in Table 4-4. The performance when using Gabor wavelets degrades compared with the results in Sections 4.3.2 and 4.3.3. Furthermore, the recognition rates of the Gabor wavelets-based methods are even lower than that of the PCA method in some cases. This is because facial expressions are formed from the local distortions of the facial feature points, which will then affect the corresponding local texture and shape properties. In this case, the Gabor representations, which abstract the textural information about the neighborhood of each pixel, are also disturbed by the local distortions caused by changes in facial expression, which results in degradation of the performance. In contrast, PCA maintains the global structure of the input, while discarding the detailed, local information. Therefore, PCA is less sensitive to local distortions, as was also discussed in Section 2.1.3.1.

When compared to other Gabor-based methods, our proposed doubly nonlinear mapping KPCA can achieve the best performance. The nonlinear mapping function Ψ_1 considers the statistical property of the input features, so that a feature with a higher probability will be more greatly emphasized. In contrast, Ψ_2, which is derived from the eigenmask, emphasizes the features from the important facial feature points. These two mapping functions can therefore enhance two different types of complementary information for face recognition, and the method that combines both Ψ_1 and Ψ_2 can provide the optimal performance.

Table 4-4 Face recognition results with different facial expressions.

Recognition Rate (%)	PCA	GW+ PCA	GW+ KPCA	GW+ $DKPCA_1$	GW+ $DKPCA_2$	GW+ DKPCA
Yale	81.3	82.7	82.7	88.0	88.0	92.0
AR	87.2	94.2	93.8	98.4	95.9	98.8
ORL	84.1	71.4	71.4	77.8	79.4	81.0

The relative performances of the different algorithms were also evaluated for faces under perspective variations. All the testing images are selected from the ORL database with the faces either rotated out of the image plane, e.g. looking to the right, left, up and down, or rotated in the image plane, clockwise or anti-clockwise. The experimental results are tabulated in Table 4-5 and show that none of these face recognition methods can achieve a satisfactory performance under perspective variations. Nevertheless, the DKPCA still outperforms other methods.

Table 4-5 Face recognition results under various perspectives.

Recognition Rate (%)	PCA	GW+ PCA	GW+ KPCA	GW+ $DKPCA_1$	GW+ $DKPCA_2$	GW+ DKPCA
ORL	48.2	56.5	57.4	64.8	60.2	66.7

4.3.5 Face Recognition with Different Databases

We have evaluated and discussed the effect of different conditions on the different face recognition methods. In this section, we also show the respective performances of the different face recognition methods based on the different databases without dividing them into sub-databases. The recognition results are tabulated in Table 4-6, which also show that the proposed doubly nonlinear mapping Gabor-based KPCA outperforms all the other methods based on the databases. In addition, our method using either Ψ_1 or Ψ_2 also outperforms the conventional Gabor-based methods in most of the cases. With these four databases, the recognition rate for the ORL database is always the lowest, irrespective of which method is used because most of the faces in this database are under perspective variations, as discussed in Section 4.3.4. Comparing the experimental results in Table 4-6 and Table 3-8, we can find that for the PCA method, the Mahalanobis distance outperforms the Euclidean distance in most cases; however, if the lighting conditions are violently uneven, such as in the YaleB database, the latter performs better.

Table 4-6 Face recognition results based on different databases.

Recognition Rate (%)	PCA	GW+ PCA	GW+ KPCA	GW+ DKPCA$_1$	GW+ DKPCA$_2$	GW+ DKPCA
Yale	80.0	87.3	88.0	91.3	90.7	94.0
AR	83.6	95.5	95.2	98.7	96.7	98.8
ORL	71.4	74.2	74.2	79.4	77.5	81.1
YaleB	65.0	93.8	95.6	98.4	94.5	98.6

4.3.6 Face Recognition with Empirical Modeling of the Feature Distribution

In Condition 1 (4.6), we assume that the probability density function (pdf) of an input feature, $p(y)$, is approximated by a normal distribution with zero mean and unit variance. Experimental results in Section 4.3.1 also show that when the parameter a (the variance of the normal distribution) is set at 1.0, the best performance can be achieved. In this section, we will discuss how this assumption satisfies the real case, and whether there are other pdf that can be used to build a more reliable mapping function.

We combine the four training sets together to form a new database, which has a total of 186 training images. All the images have a frontal view, with normal illumination and neutral expressions. Gabor wavelets are used to abstract the input features, which are then normalized to have zero mean and unit variance. Then, the pdf of y is represented by a series of discrete values, which can be computed by

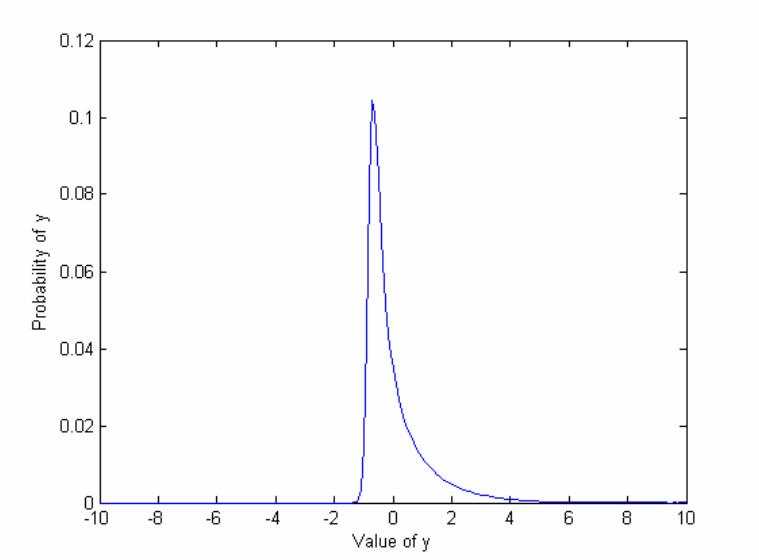
$$p_p(y) = \sum_{k=-\infty}^{+\infty} \frac{n_{kT}}{n_{total}} \cdot \delta(y - kT), \tag{4.17}$$

where $\delta(m) = \begin{cases} 1 & \text{if } m = 0 \\ 0 & \text{if } m \neq 0 \end{cases}$, n_{kT} is the number of features within the range [kT, (k+1)T], n_{total} is the total number of input features, and T is the interval, which is set at 0.1 in our experiment. Considering (4.6) and (4.17), we have

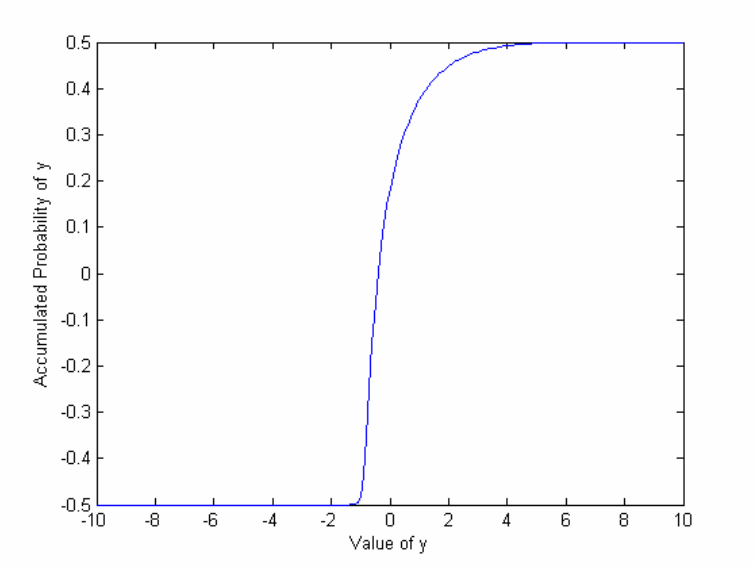
$$\Psi_{1p}(y) = \int_{-\infty}^{y} p_p(y)\,dy + \xi = \sum_{k=-\infty}^{+\infty} \left(\sum_{t=-\infty}^{k} \frac{n_{tT}}{n_{total}} \right) \cdot \delta(y - kT) + \xi, \tag{4.18}$$

where ξ is a constant used to center the mapped data. As $\sum_{k=-\infty}^{+\infty} \frac{n_{kT}}{n_{total}} = 1$, so ξ is set at -0.5. $\Psi_{1p}(y)$ is represented as a sequence of discrete values. For an input *y*, the value of $\Psi_{1p}(y)$ is computed by a linear interpolation method. Figure 4-4 shows the graphs of p_p and Ψ_{1p}.

From Figure 4-4(a), we can see that the real distribution of *y* is close to a normal distribution; however, it is not symmetrical. Compare Figure 4-4(b) and Figure 4-2(a), the graphs are also similar. We substitute this estimated Ψ_{1p} for Ψ_1 into (4.3), then repeat the procedures in Section 4.3.5. The experimental results are tabulated in Table 4-7. We can see that the recognition rates are slightly better than the results shown in Table 4-6. Considering this more simply, we can still use a normal distribution to estimate the pdf of input, which also achieves a satisfied result.



(a)



(b)

Figure 4-4 The graphs of the functions (a) $p_p(y)$ and (b) $\Psi_{1p}(y)$.

Table 4-7 Face recognition results based on different databases.

Recognition Rate (%)	Yale	AR	ORL	YaleB
GW+DKPCA	94.7	98.8	82.8	98.8

4.4 Conclusions

In this chapter, we have proposed a novel doubly nonlinear mapping Gabor-based KPCA for human face recognition. In our approach, the Gabor wavelets are used to extract facial features, then a doubly nonlinear mapping KPCA is proposed to perform feature transformation and face recognition. Compared with the conventional KPCA, an additional nonlinearly mapping is performed in the original space. Our new nonlinear mapping not only considers the statistical property of the input features, but also adopts an eigenmask to emphasize those features derived from the important facial feature points. Therefore, after the mappings, the

transformed features have a higher discriminant power, and the importance of the features adapts to the spatial importance of the face image.

This chapter has also evaluated the performances of the different face recognition algorithms in terms of changes in facial expressions, uneven illuminations, and perspective variations. Experiments were conducted based on different databases and show that our algorithm always outperforms the other algorithms, such as PCA, Gabor wavelets plus PCA, Gabor wavelets plus kernel PCA with FPP models, under different conditions. Furthermore, only one image per person is used for training in our experiments, which makes it useful for practical face recognition applications. With our approach, the recognition rates based on the Yale database, the AR database, the ORL database and the YaleB database are 94.7%, 98.8%, 82.8% and 98.8%, respectively. These results always outperform the results shown in Table 3-8, which is based on the ESTM method; this is because the former can encode higher order statistics and the features are recoded according to their statistical property and shape importance.

In Chapters 3 and 4, we describe two different face recognition techniques. We observe that, although they are robust to lighting conditions to a certain extent, their performances will also degrade when the lighting conditions are poor. To achieve good performance under poor lighting conditions, it is necessary to have pre-processing techniques for modeling the lighting to reduce or compensate for the effect. In the next two chapters, we will present two different approaches to tackling the lighting problems.

Chapter 5. Face Recognition under Varying Illumination Based on a 2D Face Shape Model

In Chapters 3 and 4, we present two methods for for face recognition under various conditions. In this chapter, we consider the case that the input image is under varying lighting conditions, and propose a novel illumination compensation algorithm, which can compensate for the uneven illuminations on human faces and reconstruct face images in normal lighting conditions based on a 2D face shape model.

5.1 Introduction

As discussed in Section 2.2, due to difficulty in controlling the lighting conditions in practical applications, variable illumination is one of the most challenging problems with face recognition. Section 2.2.2 has reviewed some methods, which are used for face recognition under varying illumination. In fact, there are also some methods, which can be considered as preprocessing algorithms before recognition. These methods are simple and computationally efficient, and also can improve the system performance to a certain extent. Histogram equalization (HE) is a commonly used method to convert an image so it has a uniform histogram, which is considered to produce an “optimal” overall contrast in the image. However, after being processed by HE, the lighting condition of an image under uneven illumination may sometimes turn to be even more uneven. Adaptive histogram

equalization (AHE) [186] computes the histogram of a local image region centered at a given pixel to determine the mapped value for that pixel; this can achieve a local contrast enhancement. However, the enhancement often leads to noise amplification in “flat” regions, and “ring” artifacts at strong edges. In addition, this technique is computationally intensive. [187, 188] introduced some modified AHE methods. Fahnestock and Schowengerdt [189] proposed a Local Range Modification, but similar problems still occur. Zhu *et al*. [190] proposed an illumination correction method, which uses an affine transformation lighting model based on a local estimation of background and the gain. However, the method is useful only when the images are under slowly varying illumination.

In this chapter, we first propose a block-based histogram equalization (BHE) method, which enhances local contrast. The locally enhanced image is then compared to a globally enhanced image, which is obtained by performing histogram equalization on the whole image; an illumination map for the face image is generated. The illumination map reflects the effect of the light source on different locations over the face image, and can therefore be used to determine the category of the light source. Based on the category, a corresponding lighting model is selected to compensate for the uneven illumination, and an image with normal lighting condition can be reconstructed. In order to correct uneven illumination without disturbing the shape information on a face image, a 2D face shape model is adopted, and all the lighting compensation is performed on the texture image. This can preserve the shape of the human face under processing.

This chapter is organized as follows. Our BHE algorithm is presented in Section 5.2. Section 5.3 describes our scheme to identify a lighting category and a new illumination correction method based on a 2D face shape model. Experimental results are given in Section 5.4, which shows the compensation results and measures the face recognition rates based on the PCA method with and without using our algorithm. Finally, conclusions are drawn in Section 5.5.

5.2 Block-based Histogram Equalization Method

A light source should have a different effect on different regions of a human face. Therefore, to determine the type of light source, one effective method is to compare the face images enhanced locally and globally. Local enhancement is described in this section, while a globally enhanced image is obtained by performing histogram equalization (HE) over the whole image. In our approach, an image is divided into a number of small blocks, and histogram equalization is performed within each of the image blocks. The pixel intensities in each image block are altered such that the resulting block has a histogram of constant intensity. In other words, all the pixel intensities within a block are modified after the BHE processing. Histogram equalization can increase the contrast in an image block, and the detailed information such as textures and edges weakened by varying illumination can be strengthened. However, this equalization process will increase the difference between the pixels at the borders of adjacent blocks.

In order to avoid the discontinuity between adjacent blocks, they are overlapped by half with each other. Weighted averaging is then applied to smooth the boundaries, i.e.

$$f(x,y)=\sum_{i=1}^{N}\omega_i(x,y)*f_i(x,y), \tag{5.1}$$

where $f_i(x, y)$ and $f(x, y)$ are the intensity values at (x, y) of block i and the smoothed image, respectively, N is the number of overlapping blocks involved in computing the value at (x, y), and $\omega_i(x, y)$, where $i = 1, \ldots, N$, is a weighting function for block i. The value of N depends on the position of the image block under consideration, which is 4 when the block is not at the border, and 2 or 1 when it is located at the border or at one of the four corners of an image. The weighting function $\omega_i(x, y)$ is simply a product of individual weighting functions in the x and y directions, i.e.

$$\omega_i(x,y)=\omega'(x)*\omega'(y), \tag{5.2}$$

where $\omega'(\cdot)$ is a triangle (hat) function, as shown below,

$$\omega'(x)=1-\left|\frac{x-S_B/2}{S_B/2}\right|, \tag{5.3}$$

where S_B is the length of a block, and x is its relative x-coordinate in the block. Thus, we have $\omega'(0)=0$, $\omega'(S_B)=0$ and $\omega'(S_B/2)=1$.

Figure 5-1 illustrates how to determine the combined intensity values of those pixels overlapped by the four adjacent blocks, i.e. block i, where $i = 1, \ldots, 4$. Histogram equalization is first performed in each of the blocks, and the combined pixel intensities can be computed by means of (5.1). BHE is simple and the computation required is much lower than that of AHE. The uneven illumination can be compensated for without the requirement of any prior knowledge, such as the direction and distribution of the light source. However, similar to AHE, noises are also enhanced after being processed by BHE. Therefore, in our approach, we only

use BHE to produce a reference image, which will then be compared to the image equalized globally.

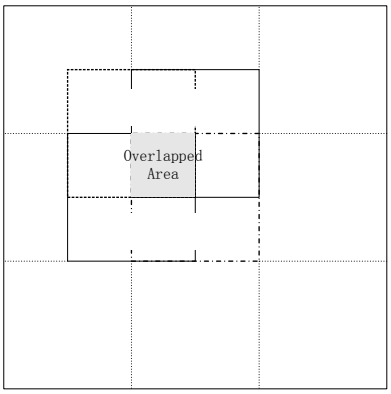


Figure 5-1 Block-based histogram equalization.

5.3 A Varying Illumination Compensation Algorithm

A face image is assumed to be a Lambertian surface, which can be described by the product of the albedo and the cosine angle between the point light source and the surface normal as follows:

$$I(x,y) = \rho(x,y)\mathbf{n}(x,y)\cdot\mathbf{s}, \tag{5.4}$$

where $I(x,y)$ is the intensity value observed of the pixel at (x, y) in the image, $0 \le \rho(x,y) \le 1$ is the corresponding albedo, $\mathbf{n}(x,y)$ is the surface normal direction, $\mathbf{s}$ is the light source direction, and its magnitude is the light source intensity.

Shashua [146] proposed that different people have the same surface normal but with different albedo, and Zhao et al. [32] also adopted this idea. However, in most natural images, albedo change is the predominant factor that causes the gradient of intensity [144], and the geometric influence cannot be neglected, especially under severe uneven lighting conditions. In these situations, the shadow

is highly dependent on the shape of a face. Therefore, in our approach, a 2D face shape model is adopted to eliminate the shape effect.

5.3.1 2D Face Shape Model

Suppose that the pixelwise correspondence between an input image and a reference face image is known, which can be determined by facial feature detection [174, 181]. The input image can be separated into texture and shape using a 2D face shape model [169]. The shape of a face is coded as the displacement field from the reference image, and the texture denotes an intensity map, which results from mapping the original image onto the reference image. All texture images have the same shape as the reference image. In our approach, uneven illumination compensation is performed on the texture image in order to avoid disturbing the shape information on the original image. After illumination compensation, the compensated texture and the original shape are combined to obtain the reconstructed image.

It is a challenge to find the pixelwise correspondence between two pictures, especially when they are under uneven lighting conditions. In our method, the position of some facial feature points, such as the eyebrows, eyes, nose and mouth, are first determined manually, as shown in Figure 5-2. Then, the displacements of these key points between a facial image and the reference image are computed. The reference shape is obtained from the average 10 size-normalized and aligned images from the YaleB database [133]. Then, using a triangle-based cubic interpolation method [191], we can map the input to the reference shape model. After processing

the mapped texture, it can be mapped backwards from the reference shape to that of the original shape.

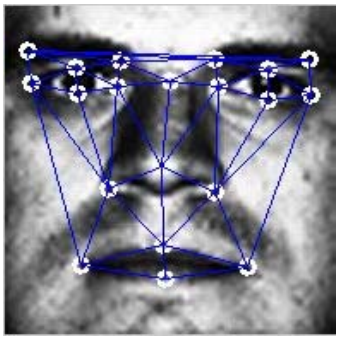


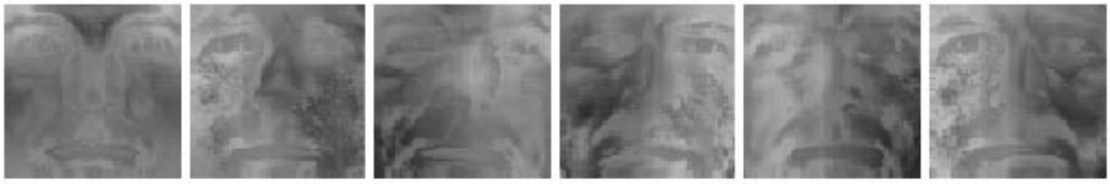
Figure 5-2 Facial feature points that are used to build a pixelwise correspondence.

5.3.2 Categories of Light Source

In our approach, the YaleB database is used as the training set, which includes 10 people; each person has 65 images with different lighting conditions. According to the illumination categories used in the YaleB database, we also divide the lighting conditions into 65 categories. Each of the categories has different azimuth angles and elevation angles of the lighting. The azimuth angles in the database vary from -130° to +130°, and the range of the elevation angle is from -40° to +90°. If both the azimuth angle and the elevation angle are equal to 0°, we say that the subject is under normal illumination.

Besides the effect of illumination on appearance, face images of distinct subjects actually look quite different. This is because the appearance of a human face is also dependent on other factors, such as gender, race, makeup, etc. Therefore, if we want to estimate the light source category, we have to eliminate the personal appearance as much as possible while keeping the illumination information unchanged. In this chapter, we use the illumination map to determine the illumination category. An image processed by BHE is considered as a reference

image. This BHE-processed image is then compared to the same image processed by HE to obtain a pixelwise difference between the two images. This difference image, which is called an illumination map, reflects the effect of the light source on different locations on the face image, and can therefore be used to estimate the illumination category. Figure 5-3 shows some examples of the illumination map with different lighting conditions.



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

Figure 5-3 Some examples of illumination map: The azimuth angles of images (a) to (f) are: 0°, 0°, 20°, 70°, -35°, -70°, respectively. The corresponding elevation angles are: 0°, 90°, -40°, 45°, -20°, 45°, respectively.

To determine the illumination category of a query image, its illumination map is first computed. Then, LDA [42] is used to determine the illumination category of the image. In order to overcome the limitation of LDA on a small sample size, we adopt the method proposed by Zhao *et al*. [192] by adding a small perturbation to all the eigenvalues such that the within-class scatter matrix S_w becomes non-singular. In our approach, the training images are divided into 65 different categories, and each category includes 9 images that are under the same lighting condition and belong to different people in the YaleB database.

5.3.3 Lighting Compensation

For each point (x, y) in an image, the effect of illumination can be written as follows:

$$f'(x,y) = A_i(x,y) \cdot f(x,y) + B_i(x,y), \qquad i = 1, \cdots, 65, \quad (5.5)$$

where $f(x, y)$ and $f'(x,y)$ represent the intensity values of the image under normal lighting condition and the image under a certain kind of illumination, respectively. $A_i(x, y)$ denotes the multiplication noise and $B_i(x, y)$ is the additive noise for the illumination mode i. The procedure deriving (5.5) from (5.4) is proved in Section 6.2.

After mapping a face image to a specific shape by the 2D face shape model and determining its illumination mode, we can compensate for the lighting effect on the face in order to generate an image with a normal lighting condition by means of the functions $A_i(x, y)$ and $B_i(x, y)$. These two functions depend on the lighting category, and we assume that they are more or less the same for images under the same illumination condition. Based on the training images in the YaleB database, we can estimate the optimal values for $A_i(x, y)$ and $B_i(x, y)$ for each illumination category by means of the least-squared method. For each illumination category i, suppose that the number of training samples equals m; then we rewrite (5.5) as follows:

$$\begin{bmatrix} f_1'(x,y) \\ \vdots \\ f_k'(x,y) \\ \vdots \\ f_m'(x,y) \end{bmatrix} = \begin{bmatrix} f_1(x,y) & 1 \\ \vdots & \vdots \\ f_k(x,y) & 1 \\ \vdots & \vdots \\ f_m(x,y) & 1 \end{bmatrix} \begin{bmatrix} A_i(x,y) \\ B_i(x,y) \end{bmatrix}, \quad \begin{matrix} i = 1, \ ..., \ 65, \\ k = 1, \ ..., \ m, \end{matrix} \qquad (5.6)$$

Let $\mathbf{F}' = \begin{bmatrix} f_1'(x,y) & ... & f_k'(x,y) & ... & f_m'(x,y) \end{bmatrix}^T$, where T represents the transpose, $f_k'(x,y)$ is the k^{th} subject under the i^{th} lighting category in the training set, and

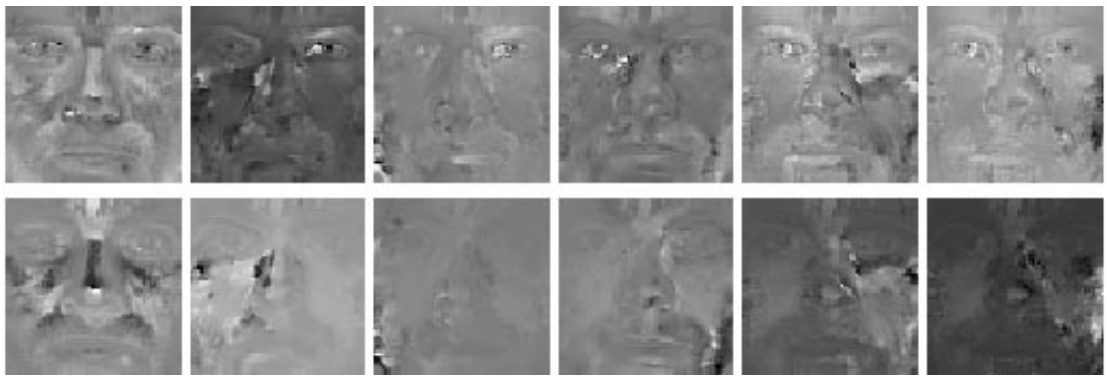
$\mathbf{F} = \begin{bmatrix} f_1(x,y) & \cdots & f_k(x,y) & \cdots & f_m(x,y) \\ 1 & \cdots & 1 & \cdots & 1 \end{bmatrix}^T$, where $f_k(x, y)$ represents a face under normal lighting condition of the k^{th} subject in the training set. Then, (5.6) can be written as follows:

$$\mathbf{F}' = \mathbf{F}\begin{bmatrix} A_i(x,y) \\ B_i(x,y) \end{bmatrix}, \qquad i = 1, \ ..., \ 65. \tag{5.7}$$

As the images in the different row of F, i.e. $f_k(x, y)$, are images of different people, they are therefore independent of each other. The least-squared solution to (5.7) can be calculated as follows:

$$\begin{bmatrix} A_i(x,y) \\ B_i(x,y) \end{bmatrix} = \left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{F}', \qquad i = 1, \ ..., \ 65. \tag{5.8}$$

Using (5.8), we can compute the optimal value of $A_i(x, y)$ and $B_i(x, y)$ for the i^{th} lighting category, and $A_i(x, y)$ and $B_i(x, y)$ are called A-map and B-map, respectively. Some examples of these two maps are shown in Figure 5-4.



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

Figure 5-4 Some examples of the A-map and B-map: A-maps are shown on the top row, and B-maps in the bottom row. The azimuth angles of images (a) to (f) are: 0°, 70°, 110°, -50°, -110°, and -130°, respectively. The corresponding elevation angles are: 20°, 45°, -20°, -40°, 40°, and 20°, respectively.

Based on the A-map and B-map of a category, the corresponding image $f(x, y)$ which is under normal lighting can be computed from $f'(x, y)$, i.e.

$$f(x,y) = \frac{f'(x,y) - B_i(x,y)}{A_i(x,y)}, \quad i = 1, \ ..., \ 65. \tag{5.9}$$

In order to avoid overflowing, all the intensity values of $f(x, y)$ are restricted to the range of [0, 255], so (5.9) is rewritten as below.

$$f(x,y) = \begin{cases} 0 & f(x,y) < 0 \\ 255 & f(x,y) > 255, \\ \dfrac{f'(x,y) - B_i(x,y)}{A_i(x,y)} & \text{otherwise} \end{cases} \qquad i = 1, ..., 65. \tag{5.10}$$

As $A_i(x, y)$ and $B_i(x, y)$ are known if the lighting category i has been determined, we can use (5.10) to construct a face image whose texture is under normal illumination. The illumination-compensated texture is then mapped from the normal shape to its original shape, and the final face image under normal lighting condition can be constructed.

5.4 Experimental Results

5.4.1 The Block Size for BHE

The block size for the BHE process will affect the performance in determining the lighting category, as well as so the performance in compensating for the illumination effect and the rate for face recognition. Table 5-1 shows the recognition rates with different block sizes used in BHE. The number of training images used is 15, and the number of test images is 150. All these images come from the Yale database [193]; they have different facial expressions and are under different illumination conditions. The PCA was used in the experiment.

Table 5-1 BHE with different block sizes.

α	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	HE
Recognition Rate (%)	52.7	62.7	66.7	60.7	62.0	60.0	60.7	56.7	54.0

The block size to be used should be proportional to the size of the face under consideration. In our scheme, the block size is set based on the distance between the two eyes of a face. The block size S_B is therefore set at α*$DisEye$, where α is a coefficient and *DisEye* denotes the distance between the two eyes. If the block size increases to the width of the whole image, BHE will be the same as HE. This result is shown in the last column of Table 5-1.

From the experimental results, a block size of 0.5**DisEye* will give the best performance level. Therefore, in the rest of the experiments, the block size for BHE is also set at 0.5**DisEye*.

5.4.2 Face Recognition Based on Different Databases

Our algorithm pre-processes a face image so that a face under uneven lighting will be converted to having even lighting. The training of our algorithm is based on the YaleB database. In this section, we will evaluate the performance of the algorithm with the use of other databases. The databases to be used include the Yale face database, YaleB face database, and AR face database [194]. The number of distinct subjects and the number of testing images in the respective databases, as well as a combination of the three databases, are tabulated in Table 5-2. For each database, only images with an upright frontal view and a neutral expression are selected. Figures 5-5, 5-6, and 5-7 illustrate the original images in the databases on the first row, those images processed by HE on the second row, those processed by BHE on the third row, and those processed by our algorithm on the fourth row.

Figure 5-5 Some experimental results based on the YaleB database: (a) Original images, (b) images processed by HE, (c) images processed by BHE, and (d) images processed by our algorithm.

Figure 5-6 Some experimental results based on the Yale database: (a) Original images, (b) images processed by HE, (c) images processed by BHE, and (d) images processed by our algorithm.

Figure 5-7 Some experimental results based on the AR database: (a) Original images, (b) images processed by HE, (c) images processed by BHE, and (d) images processed by our algorithm.

Table 5-2 The test databases used in the experiments.

	YaleB	Yale	AR	Combined
Subject	10	15	121	146
Testing set	640	30	363	1033

In order to evaluate the effectiveness of our algorithm on face recognition, PCA is used in our experiments to measure the recognition rates after processing the images using the different illumination compensation techniques. Yambor et al. [185] reported that a standard PCA classifier performed better when the Mahalanobis distance was used. Therefore, in our experiments, the Mahalanobis distance is also selected as the distance measure. The Mahalanobis distance is formally defined in [195], and Yambor et al. [185] gave a simplification, which is used here as follows:

$$d(\mathbf{x},\mathbf{y}) = -\sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}} x_i y_i, \tag{5.11}$$

where λ_i is the i^{th} eigenvalue corresponding to the i^{th} eigenvector, x_i and y_i are the i^{th} parameters of the vector $\mathbf{x}$ and $\mathbf{y}$, respectively.

In each database, one image for each subject with normal illumination was selected as a training sample, and others form the testing set. All images are cropped to a size of 64×64 and the locations of the two eyes are fixed. The number of eigenfaces used for the YaleB database, Yale database, AR database and the combined database are 9, 14, 120, and 145, respectively. The respective recognition rates based on the different databases are shown in Table 5-3.

Table 5-3 Face recognition results using deferent preprocessing methods.

Recognition Rate (%)	None	HE	BHE	New Method
YaleB	43.4	61.4	77.5	99.5
Yale	36.7	36.7	80.0	90.0
AR	25.9	37.7	71.3	81.8
Combined	30.1	32.2	60.0	92.7

In order to compare the recognition performances of different databases, we have also generated a common set of eigenfaces to test the performance based on different databases. In this experiment, we randomly selected 74 training samples from the three databases, 5 samples from YaleB, 8 samples from Yale, and 61 from AR, which produced 73 eigenfaces. Table 5-4 tabulates the recognition rates when these 73 eigenfaces are used.

Table 5-4 Face recognition results using PCA method with common eigenfaces.

Recognition Rate (%)	None	HE	BHE	New Method
YaleB	47.2	67.2	76.1	96.4
Yale	43.3	43.3	70.0	86.7
AR	22.3	40.5	56.7	73.6
Combined	24.1	25.7	48.5	86.0

Determination of the illumination category of an input face image is a very important procedure. In our method, we use the illumination maps (IMs) to estimate the category of a light source. As a comparison, some experiments which use the image processed by HE to estimate the illumination category are executed, and the corresponding recognition rates are tabulated in Table 5-5. For each database, its own eigenfaces and the common eigenfaces are both used.

Table 5-5 Face recognition results using different methods to determine illumination categories.

Recognition Rate (%)	Using Respective Eigenfaces		Using Common Eigenfaces	
	Using HE	Using IM	Using HE	Using IM
YaleB	99.2	99.5	93.8	96.4
Yale	63.3	90.0	83.3	86.7
AR	79.1	81.8	71.1	73.6
Combined	90.4	92.7	84.5	86.0

From the experimental results, we can conclude that:

1. When the testing image set includes images under varying illumination, using HE can improve the recognition performance as compared to that without using any pre-processing procedure. However, the improvement is very small in some cases, e.g. for the Yale database.

2. Using BHE or our new algorithm can improve the recognition rates significantly. The improvement using BHE is from 29.9% to 45.4%, and from 53.3% to 62.6% when our algorithm is used. In other words, these two methods are both effective in eliminating the effect of uneven illumination on face recognition. In addition, our new algorithm can achieve the best performance level of all the methods used in the experiment.
3. The BHE method is very simple and does not need any prior knowledge. Comparing to the traditional local contrast enhancement methods [186-188], its computational burden is much lower. The main reason for this is that all the pixels within a block are equalized in the process, rather than just a single pixel in the adaptive block enhancement method. Nevertheless, similar to the traditional local contrast enhancement methods, noise is also amplified after this process.
4. If we use the images processed by HE to estimate the illumination category, the corresponding recognition rates using the different databases will be lowered when compared to using the IM algorithm. This is because the variations between the images are affected not only by the illumination, but also other factors, such as age, gender, race, make-up, etc. The illumination map can eliminate the personal information as much as possible, while keeping the illumination information unchanged. Therefore, the illumination category can be estimated more accurately, and a more suitable illumination mode is selected.

5. The reconstructed facial images using our algorithm appear to be very natural, and can produce a great visual improvement and lighting smoothness. The effect of uneven lighting is almost eliminated, including shadows. However, if there are glasses or a mustache, which are not Lambertian surface, in an image, some side-effects may occur under some special light source models. For instance, glasses may disappear or a mustache can be weakened.

Zhao, *et al*. [32] used the illumination ratio image to synthesize and recognize face image under varying illuminations. Their recognition error rate, based on the YaleB database, was reported as 6.7%, while ours is 0.5%. Our algorithm uses the original images as training images, while in Zhao's method, one original image and 44 synthesized images per person were used.

5.5 Conclusions

In this chapter, we propose a new algorithm which can compensate for uneven illumination over face images. In our approach, we divide the lighting models into 65 categories. An image processed by BHE is used as a reference, and is compared to the image processed by HE to estimate the lighting category. Then, the corresponding lighting model is used to compensate for the uneven illumination. All these procedures are based on a 2D face shape model.

This approach is not only useful for face recognition when the faces concerned are under varying illumination, but can also serve for face reconstruction. More importantly, the images of a query input are not required for training. In our algorithm, the 2D face shape model is adopted in order to tackle the effect of different geometries or shapes of human faces. Therefore, a more reliable and exact

reconstruction of a human face is possible, and the reconstructed face will be under normal illumination and will appear more natural visually. Experimental results also show that preprocessing the faces using our algorithm will greatly improve the recognition rate.

The major disadvantage of the algorithm proposed in this chapter is that facial feature points must be located. Under poor lighting conditions, the detection is very difficult. Therefore, in the next chapter, we will propose a simple method to reduce the effect of uneven lighting on face recognition by means of local normalization.

Chapter 6. An Efficient Illumination Normalization Method for Face Recognition

In Chapter 5, we have proposed a method to compensate for the uneven illumination based on a 2D shape model. However, when the lighting is uneven, it is difficult to detect the position of the feature points, and to construct an accurate shape-free texture. In this chapter, we propose an efficient and effective illumination normalization algorithm, which need not perform any shape normalization and is totally automatic.

6.1 Introduction

In this chapter, a novel illumination normalization method for human face recognition is proposed. In our method, a human face is treated as a combination of a sequence of small and flat facets. The effect of the illumination on each facet is modeled by a multiplicative noise and an additive noise. Therefore, a local normalization (LN) technique [196] is applied to the image, which can effectively and efficiently eliminate the effect of uneven illumination. Then the generated images, which are insensitive to illumination variations, are used for face recognition using different methods, such as PCA, ICA and Gabor wavelets.

This chapter is organized as follows. In Section 6.2, the human face and illumination models adopted in this chapter are introduced. The LN method, which is used to eliminate the effect of uneven illuminations, is presented in Section 6.3. In Section 6.4, experimental results are detailed and the use of different illumination

compensation/normalization algorithms with different face recognition algorithms based on different databases are evaluated. Finally, in Section 6.5, conclusions are drawn.

6.2 Human Face Model and Illumination Model

As discussed in Section 5.3, a face image is supposed to be a Lambertian surface, which can be described as the product of the albedo and the cosine angle between the point light source and the surface normal as follows:

$$I(x,y) = \rho(x,y)\mathbf{n}(x,y)\cdot\mathbf{s}, \tag{6.1}$$

where $I(x,y)$ is the intensity value of the pixel at (x,y) in the image, $0 \le \rho(x,y) \le 1$ is the corresponding albedo, $\mathbf{n}(x,y)$ is the surface normal direction, $\mathbf{s}$ is the light source direction, and its magnitude is the light source intensity.

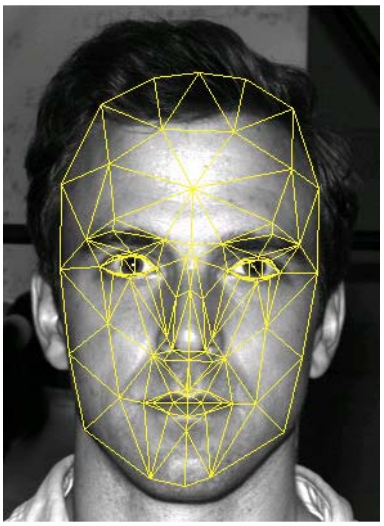


Figure 6-1 A human face image and its corresponding CANDIDE-3 model.

In computer graphics applications, a human face is treated as a combination of a sequence of small and flat facets [197, 170], which can be determined by

important facial feature points. Figure 6-1 shows a face image overlaid with an updated version of the CANDIDE model [198], which is composed of a sequence of triangular facets.

The area of each facet W is small enough to be considered a planar patch. Therefore, for each point $(x, y) \in W$, the surface normal direction $\mathbf{n}(x, y)$ is a constant. Furthermore, we assume that the light source used is directional, and therefore a good approximation of real situations [142]. Thus, the light source direction $\mathbf{s}$ is almost constant within W. Then, from (6.1), it is clear that the intensity value of the pixel at (x, y) is equal to the multiplication of the albedo at (x, y) and a scalar, which is constant within W. Suppose $f(x, y)$ and $f'(x, y)$ represent the pixel intensity values at (x, y) of the image under normal lighting conditions and the image under a certain kind of illumination, and $\mathbf{s}$ and $\mathbf{s}'$ are the corresponding light source directions. Then the corresponding illumination ratio image [32] is given as follows:

$$\begin{aligned} R_i &= f'(x, y) / f(x, y) \\ &= \left(\rho(x, y)\mathbf{n}(x, y) \cdot \mathbf{s}'\right) / \left(\rho(x, y)\mathbf{n}(x, y) \cdot \mathbf{s}\right) \\ &= \left(\mathbf{n}(x, y) \cdot \mathbf{s}'\right) / \left(\mathbf{n}(x, y) \cdot \mathbf{s}\right) = A, \quad (x, y) \in W, \end{aligned} \tag{6.2}$$

where A is determined by the surface normal direction $\mathbf{n}$ of W and the kind of illumination concerned. For a special kind of illumination, the value of A is fixed within the facet W. From (6.2), we can obtain:

$$f'(x, y) = A \cdot f(x, y), \qquad (x, y) \in W. \tag{6.3}$$

If we consider the effect of noise at each point $(x, y) \in W$, the illumination model in (6.3) can be extended to the following:

$$f'(x,y) = A \cdot f(x,y) + B, \qquad (x,y) \in W, \tag{6.4}$$

where A and B denote the multiplicative noise and the additive noise for the pixel (x,y), respectively, and they are constant within W. In (6.4), $f'(x,y)$ is the intensity value at (x,y). A and B are unknown, and the problem is how, given $f'(x,y)$, to estimate the intensity value $f(x,y)$ of the face image under normal illumination. This is an ill-posed problem. Although we assume that the values of A and B are constant in a facet W, the real range of W is unknown as it depends on the shape of a face image and is difficult to obtain under varying illumination. In Chapter 5, a 2D face shape model is adopted to map an image into a shape-free texture, and the YaleB database was then used to form the training set to obtain the A and B values pixel by pixel for each lighting category (A and B are called A-map and B-map, respectively, in this case). In this chapter, instead of estimating the values of A and B, we eliminate the effect of A and B by using the local normalization technique.

6.3 Local Normalization Technique

The main idea behind the LN technique is that, after processing an image $f'(x,y)$, its intensity value $f'_P(x,y)$ is of local zero mean and with unit variance within a facet W, i.e.

$$E\left(f'_P(x,y)\right) = 0 \text{ and } Var\left(f'_P(x,y)\right) = 1, \tag{6.5}$$

where $(x,y) \in W$. We define

$$f'_P(x,y) = \frac{f'(x,y) - E\left(f'(x,y)\right)}{Var\left(f'(x,y)\right)}, \qquad (x,y) \in W, \tag{6.6}$$

where $E\left(f'(x,y)\right)$ is the mean of $f'(x,y)$ within W and $Var\left(f'(x,y)\right)$ is the corresponding variance. Then, from (6.4), we have

$$\begin{aligned} Var\left(f'(x,y)\right) &= \sqrt{\frac{\sum\left(f'(x,y)-E\left(f'(x,y)\right)\right)^2}{N}} \\ &= A\cdot\sqrt{\frac{\sum\left(f(x,y)-E\left(f(x,y)\right)\right)^2}{N}} \\ &= A\cdot Var\left(f(x,y)\right), \qquad (x,y)\in W, \end{aligned} \tag{6.7}$$

and

$$\begin{aligned} E\left(f'(x,y)\right) &= E\left(A\cdot f(x,y)+B\right) \\ &= A\cdot E\left(f(x,y)\right)+B, \qquad (x,y)\in W, \end{aligned} \tag{6.8}$$

where N is the number of pixels within W, $E\left(f(x,y)\right)$ and $Var\left(f(x,y)\right)$ are the corresponding local mean and local variance of $f(x,y)$. From (6.4), (6.6) – (6.8), we have

$$f_P'(x,y)=\frac{f(x,y)-E\left(f(x,y)\right)}{Var\left(f(x,y)\right)}, \qquad (x,y)\in W. \tag{6.9}$$

In order to avoid overflow, a small constant (equal to 0.01) is added to all the variance values, which does not affect the derivation of (6.9). The image $f_P'(x,y)$ satisfies the conditions in (6.5), as proved in (6.10) and (6.11), i.e.

$$\begin{aligned} E\left(f_P'(x,y)\right) &= E\left(\frac{f'(x,y)-E\left(f'(x,y)\right)}{Var\left(f'(x,y)\right)}\right) \\ &= \frac{E\left(f'(x,y)\right)-E\left(f'(x,y)\right)}{Var\left(f'(x,y)\right)}=0, \qquad (x,y)\in W, \end{aligned} \tag{6.10}$$

and

$$
\begin{aligned}
Var\left(f'_P(x,y)\right) &= \sqrt{\frac{\sum_{i=1}^{N}\left(f'_P(x_i,y_i)-E\left(f'_P(x,y)\right)\right)^2}{N}} \\
&= \sqrt{\frac{\sum_{i=1}^{N}\left(f'_P(x_i,y_i)\right)^2}{N}} \\
&= \sqrt{\frac{\sum_{i=1}^{N}\left(\frac{f'(x_i,y_i)-E\left(f'(x,y)\right)}{Var\left(f'(x,y)\right)}\right)^2}{N}}, (x,y)\in W. \\
&= \frac{\sqrt{\frac{\sum_{i=1}^{N}\left(f'(x_i,y_i)-E\left(f'(x,y)\right)\right)^2}{N}}}{Var\left(f'(x,y)\right)} \\
&= \frac{Var\left(f'(x,y)\right)}{Var\left(f'(x,y)\right)} \\
&= 1
\end{aligned}
\tag{6.11}
$$

Furthermore, it is obvious that after the LN processing, the intensity value of the pixel at (x,y) is determined only by the corresponding intensity value $f(x,y)$ of the image, which is under normal illumination, and the local statistical properties of $f(x,y)$. In other words, the effects of the uneven illumination, namely the multiplicative noise A and the additive noise B, can be eliminated completely.

As with an image $f(x,y)$ under normal illumination, after the local normalization, we have

$$
f_P(x,y) = \frac{f(x,y)-E\left(f(x,y)\right)}{Var\left(f(x,y)\right)}, \qquad (x,y)\in W. \tag{6.12}
$$

From (6.9) and (6.12), we can obtain that

$$
f'_P(x,y) = f_P(x,y), \quad (x,y)\in W. \tag{6.13}
$$

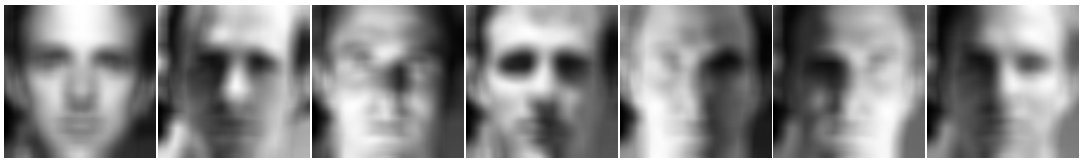
This means that, after the LN processing, the image under varying illumination will have the same intensity values as the image under normal lighting conditions. This property is very useful, and we can use the images, after LN processing, for face recognition.

Our discussion in this chapter is based on the assumption that a human face can be considered a combination of a sequence of small and flat facets. Within each facet, applying the LN technique can obtain the illumination insensitive property for each pixel. However, it is difficult to determine the range or size of a facet, especially for images under varying illuminations. In our method, we simply apply a filter of size $N \times N$ to each pixel. In other words, the filter is centered on the pixel under consideration and the corresponding mean and variance of the pixel intensities within the window are computed, then (6.6) is applied to normalize the intensity of the pixel. This process is repeated pixel by pixel to obtain a representation that is insensitive to lighting.

In (6.6), the local mean and variance of an image are computed point by point. The images formed by the local means and variances, denoted as $E\left(f\left(x,y\right)\right)$ and $Var\left(f\left(x,y\right)\right)$, are called the local mean and variance maps, respectively. Figure 6-2 illustrates some original images in the YaleB database in the first row, those images processed by histogram equalization (HE) in the second row, the corresponding local mean maps and local variance maps in the third and fourth rows, respectively, and those processed by our LN algorithm in the last row. For Figures 6-2(c) – (e), the block size used is 7×7 for local normalization.

(a) Original images.

(b) Images processed using histogram equalization.



(c) Local mean maps.

(d) Local variance maps.

(e) Images processed using LN.

Figure 6-2 Samples of cropped faces used in our experiments. The azimuth angles of the lighting of images from left to right column are: 0°, 0°, 20°, 35°, 70°, -50° and -70°, respectively. The corresponding elevation angles are: 20°, 90°, -40°, 65°, -35°, -40° and 45°, respectively.

Figure 6-2 shows that the local mean map of an image represents its low-frequency contents, while the local variance map carries the high-frequency components, or more accurately, the edge information about the image. This is because those pixels that lie in edge areas should have higher local variance values, and vice versa. In the case of uneven lighting conditions, the local mean maps are

dominated by the varying illuminations, and the edge information is disturbed by the varying local contrast and shadows. Therefore, from (6.6), we can see that in the local normalization process, the subtraction of an image by its local mean map can reduce the global uneven lighting effect, and then dividing it by its local variance map can further reduce the effect of unreliable edge information. In other words, after these two procedures, the effects of uneven illumination on both the low-frequency and high-frequency components of an image will be reduced or even eliminated. The processed image becomes robust to illumination variation and can therefore be used to achieve a more reliable performance for face recognition.

6.4 Experimental Results

In this section, we will evaluate the performance of the LN algorithm for face recognition based on different face databases. The databases used include the Yale database, the AR database, the YaleB database and the PIE database [199]. We have also combined the four databases in the experiments. The number of distinct subjects and the total number of testing images in the respective databases are tabulated in Table 6-1.

Table 6-1 The test databases used in the experiments.

	Yale	AR	YaleB	PIE	Combined
Subject	15	121	10	68	214
Testing set	30	363	640	1564	2597

For each database, the lighting conditions are different. In the Yale database, the lighting is either from the left or the right of the face images. In the AR database, besides the lighting from the left and the right, there is also lighting from

both sides of a face. The YaleB database, which consists of 10 people with 65 images of each person under different lighting conditions, is often used to investigate the effect of lighting on face recognition. In the PIE database, 24 different illumination models are adopted.

All images are cropped and normalized to a size of 64×64, and are aligned based on the two eyes. In our system, the position of the two eyes can be located either manually or automatically [173, 181], and the input color images are converted to gray-scale ones. Our method is based on the local statistical properties of images. Therefore, in order to reduce the effect of pepper noise, a 3×3 filter is adopted to detect any isolated noise point, whose intensity value will then be replaced by the mean value of the pixels within its 3×3 neighborhood.

6.4.1 The Block Size for Local Normalization

The block size used in the LN process will affect the performance in compensating for the illumination effect and, thus, the rate for face recognition. Figure 6-3 shows some images processed using the LN method with different block sizes. When the block size is very small, the statistical parameters $E(f(x, y))$ and $Var(f(x, y))$ at (x, y) are not reliable, and the output images will be noisy. However, if the block size is too large, the assumption that all the pixels within a block are located within a facet is no longer tenable, and the illumination insensitive property of the processed images also becomes invalid. Therefore, an appropriate block size is important for LN processing.

(a) The azimuth angle is 0° and the elevation angle is 20°.

(b) The azimuth angle is -50° and the elevation angle is -40°.

Figure 6-3 Face images processed using the LN technique with different block sizes. The first column shows the original images. The block sizes of other images range from 3 to 13 in increments of 2, from the left to the right column, respectively.

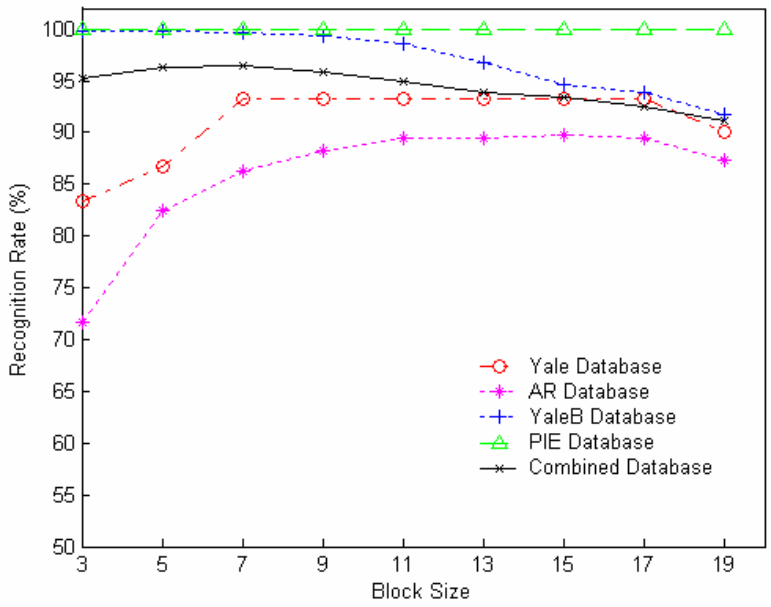


Figure 6-4 Face recognition with different block sizes.

In order to select a proper block size, PCA is used for face recognition with images processed using the LN method with different block sizes. In order to enhance the global contrast on the input images, histogram equalization is also adopted for image preprocessing (Section 6.4.2 will provide a more detailed discussion of the effect of histogram equalization). In other words, all images are

first processed by histogram equalization and local normalization sequentially, and are then followed by feature extraction and face recognition using the PCA method. Figure 6-4 shows the recognition rates based on different databases. For each database, with an increase of the block size, the recognition rate will rapidly increase until the block size reaches a critical value. Then, the recognition rate will decrease slowly. The critical or optimal filter size varies for different databases; each database has distinct characteristics in terms of the lighting conditions. We can see that the Yale and AR databases are more sensitive to the block size compared to the other databases, and the PIE database is almost independent of the window size. In our algorithm, we set the block size at 7×7, at which the combined database can obtain the best recognition rate.

6.4.2 Face Recognition Based on Different Databases

In this section, we will evaluate the performances of different lighting compensation/normalization methods for different face recognition techniques such as PCA, ICA and Gabor wavelets. The lighting compensation/normalization schemes evaluated in the experiments include the histogram equalization (HE) method, our proposed local normalization (LN) method, and the use of both HE and LN, i.e. HE+LN. We use the databases shown in Table 6-1 for testing. In each database, one frontal image of each subject with normal illumination and neutral expression was selected as a training sample, and others form the testing set.

6.4.2.1 Face Recognition Using PCA

In order to compare the recognition performances using the different databases, we used the combined database as the training set to generate a common

set of eigenfaces, which are then used for image transformation and feature extraction. The number of eigenfaces used is 213. The Mahalanobis distance metric, which is a more suitable distance measure than the Euclidean distance metric for a standard PCA classifier [185, 46], is employed, and the nearest neighbor rule is then used to classify the face images. The experimental results are shown in Table 6-2. In the second row of Table 6-2, "None" means without using any preprocessing method to normalize/compensate the varying illuminations, and directly applying PCA for face recognition.

Table 6-2 Face recognition results based on different databases using PCA.

(%)	Yale	AR	YaleB	PIE	Combined
None	43.3	78.0	60.3	88.6	60.8
HE	50.0	81.0	63.3	96.8	68.4
LN	93.3	86.0	99.5	100.0	96.4
HE+LN	93.3	86.2	99.7	100.0	96.5

Tables 6-2 shows that, with the different databases, our algorithm can achieve a better performance level than if no compensation/normalization scheme is used or if only the histogram equalization is used. The performance will slightly improve when the histogram equalization is used with the local normalization method; this shows that the global contrast enhancement can improve illumination compensation to a certain extent. Comparing with the case that without using any preprocessing method, the error rate using HE plus LN based on the combined database can be reduced by 91.1%, i.e. from 39.2% to 3.5%.

As the YaleB database is commonly used to evaluate the performance of illumination invariant face recognition, so we first compare our performance with other face recognition methods based on this database. Georghiades *et al*. [138]

proposed the individual illumination cone model and achieved 100% recognition rates, but the method requires seven images of each person to obtain the shape and albedo of a face. Lee *et al*. [200] used a nine-point light source method to achieve a 99.1% recognition rate. However, the approach requires nine simulated images with different illumination variations for each person. Zhao *et al*. [32] synthesized 45 images per person, which are adopted for training, and a 93.3% recognition rate was achieved. Liu *et al*. [147] reported a 98.4% recognition rate. However, the iterative algorithm, which is used to restore the input image, is more computational than our method. All the above methods only consider the situation where the light source directions are within 75º, and so only 45 illumination models were used for testing. However, in our experiment, a total of 65 lighting conditions were tested. In our method proposed in Chapter 5, the recognition rates are 99.5% and 96.4% when the respective eigenfaces and common eigenfaces are adopted for PCA method, respectively. The results are similar to those proposed in this chapter. However, our previous method requires twenty feature points per image to determine the 2D shape of the input and to construct a shape-free texture, which is very difficult when the image is under varying or poor illumination. In [147], the recognition rate with the Yale database is reported to be 81.7%. The method proposed in Chapter 5 has also been tested based on the Yale database and AR database, and the results are 90.0% and 81.8%, respectively, when the respective eigenfaces are used, and 86.7% and 73.6%, respectively, when the common eigenfaces are adopted.

Compared to other methods, our proposed algorithm is much simpler. We neither require multiple images with different illumination variations as training, nor

require the detection of important facial feature points to perform shape normalization. Our method is robust to illumination conditions and is computationally simple, which is important as a preprocessing method. Therefore, our method can also be used for other face recognition methods.

6.4.2.2 Face Recognition Using ICA

In this chapter, we employed the FastICA [201] to compute the ICs of a set of training images. FastICA provides rapid convergence and estimates the ICs by maximizing a measure of independence among the estimated original components [45, 46]. The results in [43, 46] show that ICA will have a better performance when the cosine similarity measure is used. Therefore, we also adopt this similarity measure, which is defined as follows:

$$d(\mathbf{u},\mathbf{v}) = \cos\left(\frac{\sum_{i=1}^{k} u_i v_i}{\sqrt{\sum_{i=1}^{k} u_i^2}\cdot\sqrt{\sum_{i=1}^{k} v_i^2}}\right), \tag{6.12}$$

where u_i and v_i represent the i^{th} element of two k-dimensional feature vectors **u** and **v**, respectively. We also use the combined database as shown in Table 6-1 to produce the ICs, the number of ICs used being 214. The experimental results are shown in Table 6-3.

Table 6-3 Face recognition results based on different databases using ICA.

(%)	Yale	AR	YaleB	PIE	Combined
None	40.0	77.4	65.6	95.1	64.8
HE	53.3	78.5	72.0	97.5	75.4
LN	83.3	82.4	98.1	100.0	90.6
HE+LN	86.7	82.6	99.8	100.0	94.5

Comparing Table 6-2 and Table 6-3, we can see that when the first two methods ('None' and HE) are used, ICA outperforms PCA in most of the cases. However, when our proposed LN method is employed with or without using the HE method, PCA outperforms ICA in most of the cases. As described in [67], uneven illuminations mainly affect the global components of a face image. Therefore, when the input image is under varying lighting conditions without any preprocessing method or when the HE method only is used for illumination normalization, ICA, which maintains more local, detailed information, performs better than PCA, which mainly considers the global structure of an input. This result coincides with the analysis in [46]. When our LN method, which can effectively enhance the local structure of an image and reduce the global effect of the varying illumination, is used, more local and detailed texture will appear in the processed image. In this case, PCA can more effectively represent the more important structure of an image and reduce the effect of the noise enhanced by local normalization. Therefore, after the LN process, PCA outperforms ICA. In fact, the difference between these two methods is not large, especially for the YaleB database and the PIE database, where both methods can achieve a recognition rate near 100% (the Yale database is an exception, but its size is very small). We have also conducted some experiments in which the Euclidean distance metric is employed. For the combined database, the recognition rate without using any illumination normalization method is 62.4%, and the results using HE, LN and HE plus LN are 68.0%, 89.1% and 93.7%, respectively. These results are lower than those shown in Table 6-3, but the relative performances of these methods remain the same.

6.4.2.3 Face Recognition Using Gabor Wavelets

Section 2.1.3.5 describes how to use the Gabor wavelets to perform face recognition. In this part, we select one center frequency, which is equal to $\pi/2$, and eight orientations from 0 to $7\pi/8$ in increments of $\pi/8$. The Euclidean distance metric is adopted and the nearest neighbor rule is used for classification. The experimental results are shown in Table 6-4.

Table 6-4 Face recognition results based on different databases using Gabor wavelets.

(%)	Yale	AR	YaleB	PIE	Combined
None	63.3	90.9	86.7	99.9	86.1
HE	73.3	94.5	98.4	100.0	90.8
LN	100.0	98.3	99.4	100.0	98.4
HE+LN	100.0	98.6	99.5	100.0	98.7

Tables 6-2 ~ 6-4 demonstrate that, of the three feature extraction methods, Gabor wavelets can achieve the best performance. Especially for the PIE database, a 99.9% recognition rate can be obtained based on the original images. This is because Gabor wavelets can effectively abstract local and discriminating features, which are less sensitive to illumination variations. It is clear that applying our LN method can further increase the performance when using Gabor wavelets for face recognition based on different databases. Liu *et al*. [147] also uses Gabor wavelets to extract features based on the restored images, and the recognition rate is 95.3% for the combined Yale database and YaleB database.

6.4.3 Computational Complexity

We have proposed an efficient method of reducing the effect of varying illumination on face recognition. Suppose that the size of a normalized face is $M \times M$,

and the block size used in the LN method is $N \times N$. The computational complexity for pre-processing an image using LN is $O(M^2N^2)$. All our experiments were conducted on a computer system with Pentium IV 2.4GHz CPU and 512MB RAM. The average runtime of our algorithm to normalize the illumination of a face image in the AR database (363 face images) is about 6.2 milliseconds, where M and N are equal to 64 and 7, respectively. As our method has a low complexity, it can also be applied to some real-time applications such as illumination normalization in video sequences.

6.5 Conclusions

In this chapter, a novel and simple illumination normalization method for human face recognition under varying lighting conditions is proposed. A human face is treated as a combination of a sequence of small and flat facets. For each facet, the effect of the illumination can be modeled by a multiplicative term and an additive term. Therefore, a local normalization technique is applied to the image point by point. Local normalization can effectively and efficiently eliminate the effect of uneven illumination, and keep the local statistical properties of the processed image the same as for the corresponding image under normal lighting conditions. Then, the generated images, which are insensitive to illumination variations, are used for face recognition, and the performances are evaluated using different face recognition methods. Experimental results show that, with the use of PCA, ICA and Gabor wavelets for face recognition, the error rates can be reduced by 91.1%, 84.4% and 90.6%, respectively, based on the combined database when our illumination normalization algorithm is used.

A major advantage of our proposed method is that, for training, only one image per person under normal illumination is required; this is very important for real applications. In addition, there is no need to perform any facial feature detection and shape normalization, which can be very complicated when the lighting is uneven or complex. Furthermore, our method is computationally simple, can serve as a preprocessing technique and also combine with other methods for face recognition. In this chapter, we only consider the situation where the human faces are frontal and have a neutral expression. For a practical face recognition application, various poses and expressions may combine with varying illuminations. If these effects are also considered, the overall recognition rates will be further improved.

Chapter 7. Facial Expression Recognition based on Shape and Texture

In Chapter 3, we propose a novel elastic shape-texture matching method, namely ESTM, for human face recognition under various conditions. In this chapter, we will apply this method for facial expression recognition. Besides ESTM, we also propose a new representation model for facial expressions, namely spatially maximum occurrence model (SMOM), which is based on the statistical characteristics of training facial images and has a powerful representation capability. Finally, ESTM and SMOM are combined together to obtain an optimal performance.

7.1 Introduction

In this chapter, a novel and accurate method is proposed for facial expression recognition. Our method includes two major techniques: spatially maximum occurrence model (SMOM), which is used to describe the different facial expressions; and elastic shape-texture matching (ESTM), which is proposed in Chapter 5 and is used to compute the similarity between two images. The combination of these two techniques, namely the SMOM-ESTM method, is used to classify the facial expressions. Due to the fact that facial expression is such a pattern whose within-class variation sometimes is larger than the between-class variation, we propose to use SMOM, which is based on the statistical properties of training images and has a powerful representation capability, instead of a sequence of fixed images to describe the expressions. The shape and texture information about a face

image are complementary and supplementary to each other, and both of which are useful for expression recognition. In [161], the line edge map, which mainly represents the shape information about a face, is used to describe an expression. Although the direction of an edge line can reflect some texture information, that information is still insufficient. Lyons *et al*. [164] adopted the 2D Gabor wavelet to describe the texture, but the feature points, which represent the shape information, have to be detected manually. ESTM can compute the similarity between two images based on both the shape and texture information, requiring only the positions of the two eyes and middle of the mouth for alignment. In our algorithm, ESTM is combined with SMOM for facial expression recognition.

This chapter is organized as follows. In Section 7.2, the principle of our proposed facial expression representation method, SMOM, is described. The ESTM for facial expression recognition is presented in Section 7.3. Then, the combination of SMOM and ESTM used for expression recognition is introduced in Section 7.4. Experimental results are given in Section 7.5, which shows the performances of our algorithms based on the AR database and the Yale database. Finally, conclusions are drawn in Section 7.6.

7.2 Spatially Maximum Occurrence Model for Representing Facial Expressions

Human facial expression is a complex pattern, which relies on the emotion of the expressor and varies from person to person. On the one hand, the expression is determined by the movements or changes in facial features, which means that it is person-dependent and is affected by the characteristics of the expressor, such as the

shapes or positions of the facial features, motion habits, and so on. On the other hand, for the same person, there are also variations in the same expression due to different degrees of emotion. Therefore, the within-class variation of an expression is relatively large, and the between-class variation of different expressions is relatively small. In fact, even human beings sometimes cannot judge the expressions correctly based on a still image. In this case, knowing how to build proper expression models is very important. Using the mean image of a training set to represent a particular expression is simple, however, most of the information is lost, and the within-class variations cannot be reflected. In this section, we will propose a new expression representation scheme, namely Spatially Maximum Occurrence Model (SMOM), which is based on the statistical properties of the training set and contains most of the significant visual content.

SMOM is constructed based on the probability of the occurrence of pixel values at each pixel position for all the training images, which is illustrated in Figure 7-1. Suppose that the number of training images is equal to N, and the size of an image is $M \times H$. Therefore, there are N possible values at each pixel position (x,y). Ranking these N intensity values, we can obtain the histogram $H_{x,y}(b)$ for the pixel position (x,y) as follows:

$$H_{x,y}(b) = \sum_{k=1}^{N} \delta(f_k(x,y) - b), \tag{7.1}$$

where $\delta(m) = \begin{cases} 1 & \text{if } m = 0 \\ 0 & \text{if } m \neq 0 \end{cases}$, for $0 \le b < B$. B is the number of bins in the histogram, and $f_k(x, y)$ is the intensity value of the k^{th} image at position (x, y). In general, B is equal to the number of intensity levels in the images. However, when the number of

training images is small, the number of bins should be reduced and the histogram should be smoothed using a Gaussian filter as follows:

$$H'_{x,y}(b) = H_{x,y}(b) * G(\sigma, b), \tag{7.2}$$

where $G(\sigma, b)$ is a Gaussian filter with variance σ, * is the convolution operator, and $H'_{x,y}(b)$ is the smoothed histogram of the pixel position (x, y). For each smoothed histogram, its peak values are identified and ranked in descending order. A peak occurs at a bin if its value is higher than its two adjacent bins. If a bin is the first (or the last) bin in a histogram, and its value is larger than the right (or the left) bin, we also consider it a peak. If m consecutive bins have the same value and this value is higher than the two adjacent bins of the consecutive bins, a peak also exists, and the bin value of the peak is set at the middle of the m consecutive bins. The gray levels corresponding to those bins that are the peaks of a histogram will be used in constructing SMOM. In other words, at each pixel position (x, y), the gray levels corresponding to the peaks are ranked according to their probabilities of occurrence. SMOM is therefore defined as follows:

$$SMOM(x, y, k) = \{b_1, b_2, \ldots, b_k\}, \tag{7.3}$$

where $0 \le b_k < B$, for $0 \le x < M$ and $0 \le y < H$, k is the number of peaks to be considered in the representation, $b_1, b_2, \ldots, b_k$ are the gray levels corresponding to the peaks of the histogram for pixel position (x, y), and the conditions $H'_{x,y}(b_1) \ge H'_{x,y}(b_2) \ge \ldots \ge H'_{x,y}(b_k)$ are satisfied. Usually, k is a small value. If the number of peaks p in a histogram is less than k, the remaining $k–p$ values will be corresponding to those bins with the largest probabilities of occurrence.

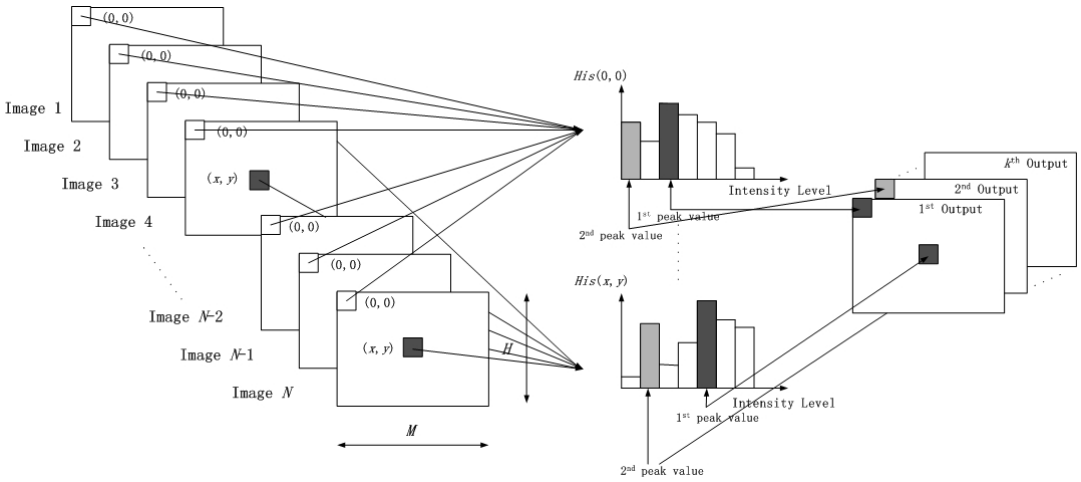


Figure 7-1 The construction of a SMOM.

In our algorithm, the gray levels of those bins corresponding to the highest peaks, rather than the highest values, are used to represent the pixel intensities. As a histogram can be considered as a multi-cluster distribution and a peak is the representation of a bin cluster, so the peak values can provide useful statistical information at a pixel position, and are suitable for modeling complex patterns, such as facial expressions. An advantage of SMOM is its powerful representation capability. Each pixel position (x, y) in SMOM is represented by k values. For an image with size $M \times H$, the number of possible images that can be generated from SMOM is K^{MH}. Suppose that $k = 2$, M and H are both equal to 64, SMOM can be used to represent $2^{64\times64} \approx 10^{1233}$ different images. Furthermore, because the representation values are based on the statistical properties of the training images, most of the significant visual content of the training set is maintained in SMOM. In our method, SMOM is used for modeling the facial expression patterns.

7.3 Elastic Shape-Texture Matching

As the statement in Chapter 3, ESTM is a method that measures the similarity between images based on their shape and texture information. The shape is represented by the edge map $E(x, y)$, and the texture is characterized by the Gabor wavelets and the gradient direction of each pixel, which are described by the Gabor map $\tilde{G}(x, y)$ and the angle map $A(x, y)$, respectively.

Nastar *et al*. [112] have investigated the relationship between variations in facial appearance and their deformation spectrum. They found that, when a facial expression varied, only the high-frequency spectrum was affected, and this is called a high-frequency phenomenon. This suggests that the high-frequency components are more discriminant for facial expressions. Therefore, in our method, we apply the Gabor wavelets on the edge images, instead of the original images, to obtain the corresponding texture information, i.e. the Gabor map, in the high-frequency spectrum.

7.4 Facial Expression Recognition

In our approach, SMOM is used to represent the different facial expressions for recognition. Suppose that there are W classes of facial expression, then W expression models, $SMOM_1$, $SMOM_2$, …, and $SMOM_W$, are constructed. For each model, k peak values are used to represent the gray-level intensity at each pixel position, and these peak values are ranked according to their probabilities of occurrence. Then, the difference between the facial expression in a query input and each of the models will be computed. As discussed in Section 7.2, a SMOM describes the possible distribution of a certain expression in the image space. The

distance from a query input to the image space generated by a SMOM is defined as follows:

$$D_m\left(f(x,y),l\right)=\sum_{x=0}^{M-1}\sum_{y=0}^{H-1}p(u')\cdot\left|f(x,y)-SMOM_l(x,y,u')\right|, \qquad (7.4)$$

where $u'=\arg\min_{u}\left(\left|f(x,y)-SMOM_l(x,y,u)\right|\right)$, for u = 1, …, k and l = 1, …, W.

$p(u')$ is a penalty function and is set as $p(u')=1/H'_{x,y}(u')$, where $H'_{x,y}(u')$ is the smoothed histogram of the pixel position (x, y), and reflects the probability of occurrence of u'. This distance measure is simple, and it computes the minimum distance between the gray-level intensities of the query and the respective peak values at each position. The smaller the value of D_m, the closer the query image is to $SMOM_l$, and the more reliable the $SMOM_l$ being used to represent the query image, and vice versa.

ESTM is adopted to perform the recognition. There are N_l training images for the expression class l, and the corresponding mean image is denoted as $\bar{f}_l(x,y)$, where $1 \le l \le W$. Each of these W mean images provides the shape and texture characteristics of the corresponding facial expression class, and is used in matching for the expression class based on the ESTM. Therefore, the shape-texture Hausdorff distance $H(A, B)$ (3.3) is used as the distance measure, which considers the similarity between images based on their shape and texture properties. Combining $H(A, B)$ with the model distance D_m, a new distance measure between the facial expressions in the query $f(x,y)$ and that of class l is defined as follows:

$$D(f(x,y),l)=\lambda\cdot\frac{D_m\left(f(x,y),l\right)}{\max_i\left(D_m\left(f(x,y),i\right)\right)}+(1-\lambda)\cdot\frac{H\left(f(x,y),\bar{f}_l(x,y)\right)}{\max_i\left(H\left(f(x,y),\bar{f}_i(x,y)\right)\right)}, \qquad (7.5)$$

where $l = 1, \ldots, W$ and $0 \leq \lambda \leq 1$. This distance measure contains two different distance measures for facial expressions, which are normalized by their respective maximum distances. λ is used to adjust the relative weights for these two terms in the distance measure. The first term is used to measure the reliability of using the l^{th} SMOM to model the input expression, while the second term provides a distance measure based on the shape and texture of the l^{th} mean image. In other words, the first term considers the statistical properties of the training set at each position and the second term uses the shape and texture information in the spatial domain. These two terms are supplementary to each other, and the combined distance measure is called the SMOM-ESTM algorithm, which can achieve a good performance in facial expression recognition.

7.5 Experimental Results

In this section, we will evaluate the performance of the SMOM-ESTM algorithm for facial expression recognition based on different face databases. The databases used include the AR database and the Yale database. All images are cropped to a size of 64×64 and are normalized to make the two eyes and the vertical position of the mouth aligned. In our system, the position of the two eyes and the middle point of the mouth can be located either manually or automatically [173, 181], and the input color images are converted to gray-scale ones. In order to enhance the global contrast of the images and to reduce the effect of uneven illuminations, histogram equalization is applied to all the images. The parameters used in (3.5) and (3.9) are set at $\alpha = 0.1$, $\beta = 0.1$, $\gamma = 0.8$, $P_e = 3$ and $P_a = \pi/10$. For ESTM, the neighborhood size used is set at 7×7, which allows the expression to

vary to a certain extent. As most of the facial muscle movements focus on the eyes (including eyebrows) and mouth area, which represent different expressions [20, 161, 163], we adopt the face model proposed in [173] to produce a facial mask, which maintains the eye and mouth areas while blocking other parts of a face. This can reduce the effect of the personal characteristics and emphasize the actions of these key features.

7.5.1 Expression Recognition based on the AR database

In the AR database, there are 121 persons, comprising 70 males and 51 females. For each person, there are three expressions: neutral, smile and scream. Figure 7-2 shows some examples from the database. We randomly selected 60 samples for each class as the training images to generate SMOM; the remaining 61 identities are used for testing. For each pixel position of the SMOM, there are k representation values. Combining the i^{th} peak values ($1 \leq i \leq k$) at each pixel position, we can construct a pattern image, which is called the i^{th} peak image. Figure 7-3 illustrates the first five peak images produced by the SMOM. The 1^{st} peak image may be considered to be the most likely pattern of the expression concerned, while the k^{th} peak image is the least likely one. These k peak images distribute evenly, to a certain extent, over the corresponding pattern space, and can therefore represent the pattern space well. The mean image of the training images of each expression class is also computed and shown in the last column in Figure 7-3. We can see that the peak images look similar to the mean image, while the mean image looks smoother. This is due to the fact that the intensity values of the k peak images are obtained based on the statistical characteristics of the training data, and this

process is dependent on the pixel position. In other words, the correlations between neighboring pixels are reduced, and the image becomes less smooth.

Figure 7-2 Some cropped faces in the AR database. Facial expressions: (a) Neutral, (b) Smile, and (c) Scream.

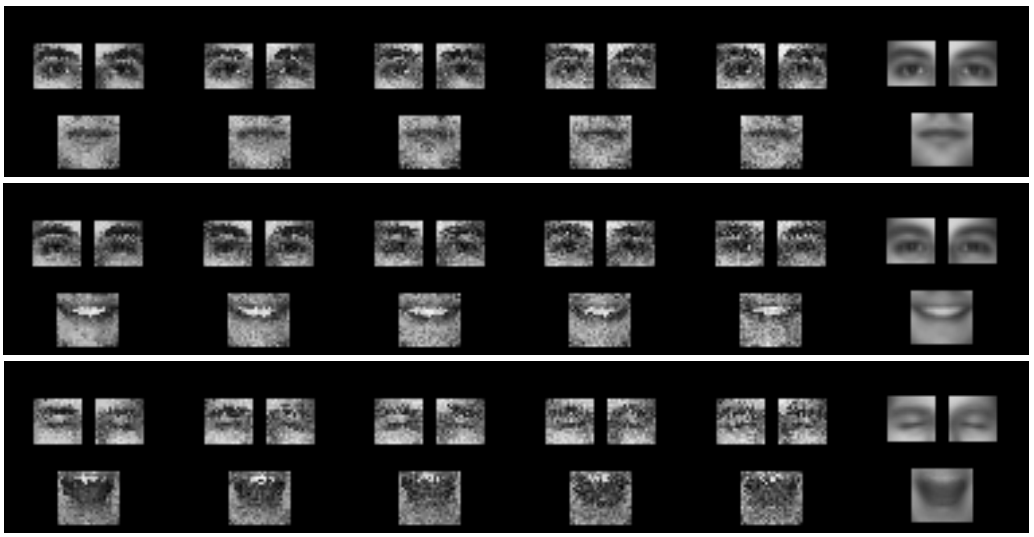


Figure 7-3 The masked mean images and peak images produced by SMOM. The right-most column displays the mean images for the expression classes, and the first to the fifth columns show the first five peak images, respectively. Facial expressions: (a) Neutral, (b) Smile, and (c) Scream.

If only SMOM is considered, (7.4) provides a distance measure to classify the input images. This is actually equivalent to setting λ to 1 in (7.5). In Section 7.4, we have used (7.4) to determine the reliability of SMOM. In fact, a more reliable SMOM means that we can recognize or classify the expressions more accurately. Therefore, (7.4) can be used for recognition directly. The experimental results are shown in the Table 7-1, where the last column shows the result when the mean images are used as the expression patterns, and the minimum distance measure is used for classification. We can see that the performance using SMOM is better than that using the mean images, even if k is equal to 1. When k increases, the recognition rate also increases until k is larger than 8. Then, the recognition rate will decrease slowly. A slight decrease in the recognition rate happens when k increases from 3 to 4; this may be caused by the perturbation of the statistical properties of the training data.

Table 7-1 Facial expression recognition rates using SMOM.

Value of k	1	2	3	4	5	6	7	8	9	10	Mean Image
Recognition Rate (%)	71.6	75.4	79.2	78.1	79.8	80.3	83.1	84.7	83.6	83.1	66.1

If only ESTM is employed for recognition, i.e. $\lambda = 0$ in (7.5), and the mean images are used for training, the result is 92.9%. The performance using ESTM is better than the case of using SMOM only, due to the latter being based on the gray-level intensities, while the former abstracts more shape and texture features.

Combining SMOM and ESTM, i.e. using (7.5) with different values of λ, we can obtain a better result than only using either one of them. Figure 7-4 shows the recognition performances of the SMOM-ESTM algorithm with different values of

λ, and the number of peaks used in SMOM is set at 8. The best performances are achieved for both the AR database and Yale database when λ is equal to 0.25. For the AR database, the highest recognition rate achieved is 94.5%. In order to compare our method with other algorithms, we also build a testing set, which includes all 121 subjects, and we then employ the SMOM-ESTM method for recognition, where the training data is unchanged. The results are shown in Table 7-2. The results reported in [161], which is also based on the AR database (only 61 males and 51 females are used), are tabulated in Table 7-3.

Table 7-2 Facial expression recognition results using the SMOM-ESTM method.

	Neutral	Smile	Scream	Average
Male	91.4	94.3	98.6	94.8
Female	100.0	98.0	98.0	98.7
Average	95.0	95.9	98.3	96.4

Table 7-3 Facial expression recognition results reported in [161].

	Neutral	Smile	Scream	Average
Male	91.8	68.9	88.5	83.1
Female	96.1	90.2	86.3	90.9
Average	93.8	78.6	87.5	86.6

Our method outperforms the method proposed in [161], which can be explained by the following: 1) SMOM, which is based on the statistical properties of the training set, can provide more reliable expression models; 2) more texture information, which is based on the Gabor wavelets and gradient direction, is abstracted and useful for describing the expression; and 3) edge points are more elastic and suitable than edge lines for expression matching, especially in the case of the existence of large shape variations. In fact, [161] also adopted a splitting process

to divide a long edge line into a sequence of short lines in order to improve its performance.

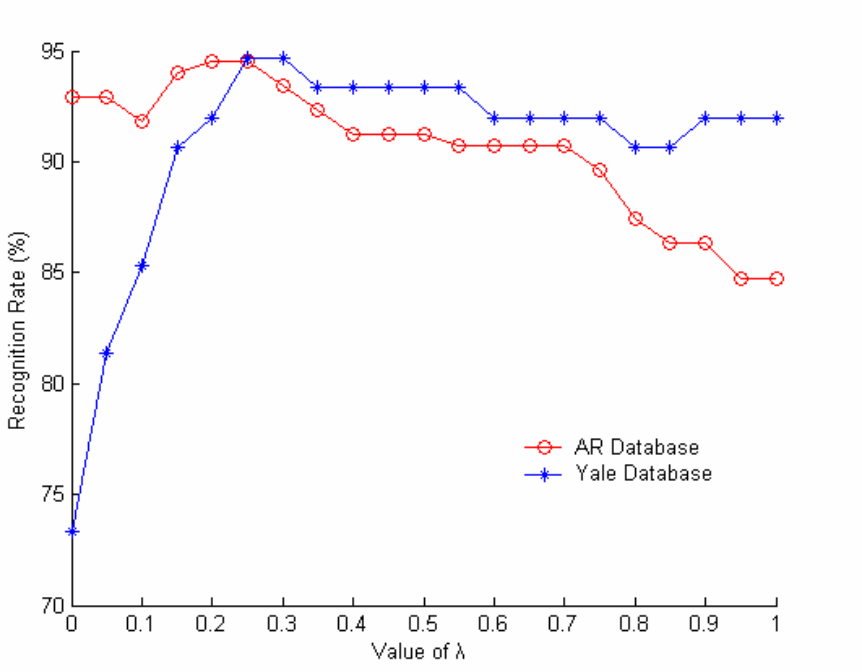


Figure 7-4 Facial-expression recognition performances of the SMOM-ESTM algorithm with different values of λ based on the AR database and the Yale database.

7.5.2 Facial Expression Recognition based on the Yale database

The Yale database includes 15 persons (14 males and 1 female). For each person, there are five expressions: neutral, smile, surprise, blink and grimace (where grimace means the left eye is closed). Some examples from the Yale database are shown in Figure 7-5. Compared with the AR database, there are far fewer subjects here but expression patterns. A leave-one-out mechanism is adopted to evaluate the recognition performances. In the experiments, all samples but one identity are used for training, and the images of that person are used for testing. This process repeats for every identity, and the results are averaged. Because SMOM is a statistical

representation of the expression classes, using 14 images for training is insufficient. By translating each original image into a pixel distance along the eight directions, we can produce eight new images. These produced images and the original image have a different type of importance when constructing the SMOM. A matrix, $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$, is therefore used to weigh the importance in computing the histograms. With this matrix, the number of pixels from the original image will be multiplied by 4, and the number of pixels from those images shifted along either the *x*-axis or the *y*-axis will be multiplied by 2 in the construction of the histograms. Thus, each identity has 16 images, and each expression class contains 224 images for training. The first five peak images produced by SMOM are shown in Figure 7-6. These peak images look similar to the corresponding mean images, which are displayed in the last column. However, due to the shift in the training images, the edges of the peak images become blurred. When computing the mean images, only the original images are used.

Figure 7-5 Some cropped faces in the Yale database. Facial expressions: (a) Neutral, (b) Smile, (c) Surprise, (d) Blink and (e) Grimace.

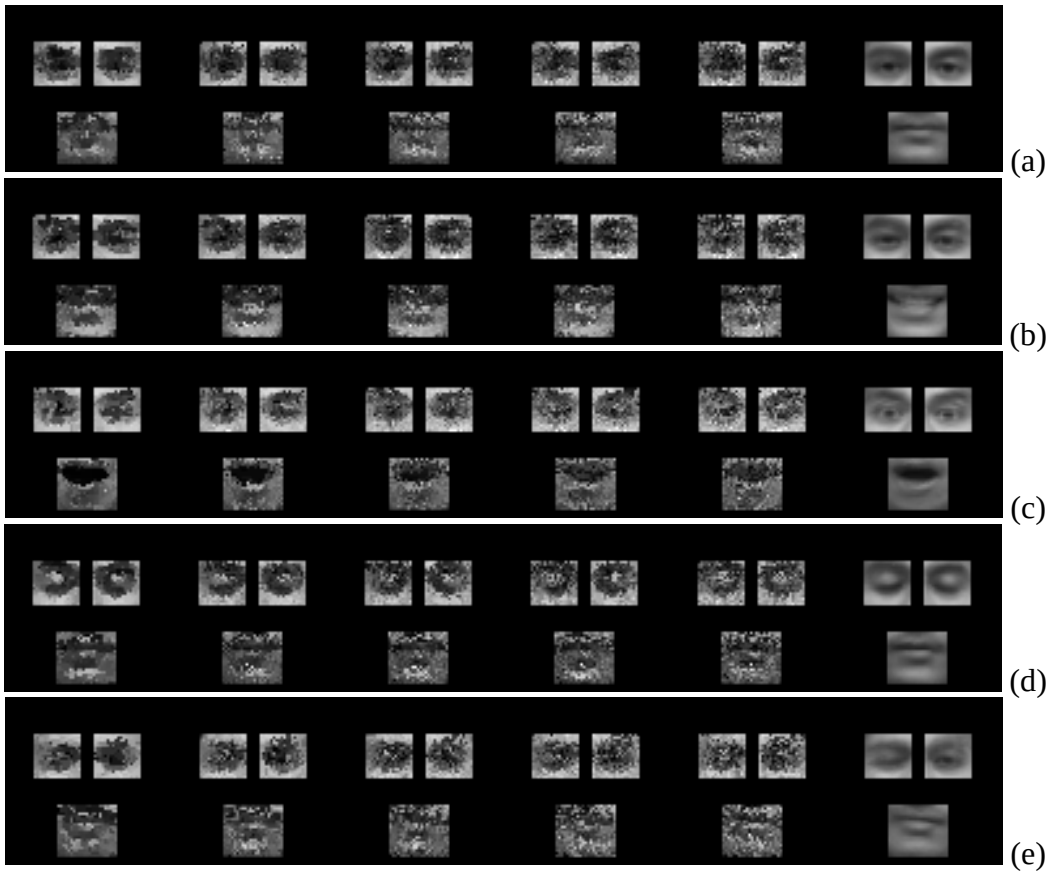


Figure 7-6 The masked mean images and peak images produced by SMOM. The right-most column displays the mean images for each expression class, the first to fifth columns show the first five peak images, respectively. Facial expressions: (a) Neutral, (b) Smile, (c) Surprise, (d) Blink and (e) Grimace.

In the following experiments, we set k at 8 for SMOM and λ at 0.25 in (7.5), as in Section 7.5.1. We evaluated the performances of the respective algorithms for different expression categories, which are tabulated in Table 7-4. For ESTM, the recognition rates are lower than those based on the AR database. This is because the expressions "Blink" and "Grimace" mainly involve movements of the eyelids. Due to the low contrast in the eye areas (even human beings cannot correctly judge the movements based on some of the original images), it is difficult to abstract valuable shape and texture information in these areas (see Figure 7-5). With the different methods, the performance for recognizing the "Smile" is always the best. This can be explained by the fact that the smile expressions of different identities are more or less the same, i.e. its within-class variation is small (unlike the expression "Surprise"). In addition, the smile expression is quite distinct from other expressions, such as "Neutral", "Blink" and "Grimace", so its between-class variation is relative large. Based on the experimental results, the SMOM-ESTM algorithm can always give the best recognition performance for different expressions.

Table 7-4 Facial expression recognition rates using different methods.

(%)	Neutral	Smile	Surprise	Blink	Grimace	Average
SMOM	86.7	100.0	93.3	86.7	93.3	92.0
ESTM	53.3	93.3	73.3	73.3	73.3	73.3
SMOM-ESTM	93.3	100.0	93.3	86.7	100.0	94.7

7.5.3 Storage Requirements and Computational Complexity

In our approach, the data stored in a database includes the expression models produced by SMOM, and by the edge maps, Gabor maps and angle maps of the

mean images for ESTM. Suppose that there are W expression models ($SMOM_1$, $SMOM_2$, …, $SMOM_W$), and for each model, k representative values are used at each pixel position with 8 bits per value. As a facial mask is adopted to emphasize the actions at the eye and mouth areas, only the points in these areas are stored. The number of pixels involved in these areas is denoted as N_S. Then, the number of bytes used for these expression models is WN_Sk. For ESTM, W mean images are used for the expression classes. The numbers of bytes for the edge map, Gabor map, and angle map of an image are $2\eta N_S$, $2n_fn_aN_S$, and $2N_S$, respectively, where η is the percentage of the points selected as edge points in an edge map, and n_f and n_a are the numbers of center frequencies and orientations used for the Gabor filters, respectively. Therefore, the total number of bytes or the storage requirement is $(k + 2\eta + 2n_fn_a + 2)WN_S$.

The computational complexity for recognizing a query face image includes two parts: feature extraction and matching. The runtime required for feature extraction is the time spent on computing the edge map, Gabor map, and angle map of the query image (SMOM performs matching based on the gray-level intensities, and does not need to extract additional features). As all the maps of the model images for the expression classes have been computed and stored in the face database, we need to consider only the time required to generate the maps of the query image. The computations required for computing an edge map, Gabor map and angle map are in the order of $O(N_S)$, $O(N_S \log_2(N_S))$ and $O(N_S)$, respectively. For searching in a large database, the runtime for matching is the most significant part of the whole process. For SMOM, the computation required for the model distance

is in the order of $O(kWN_S)$. For ESTM, the computational complexity is in the order of $O(2\eta N_S D^2 W t_{all})$, where D is the size of the neighborhood considered when performing the elastic matching, and $t_{all} = t_e + t_g + t_a$, where t_e, t_g, t_a are the average runtimes required to compute the edge distance, Gabor distance and angle distance, respectively, for a point pair. Therefore, the total computational complexity for recognizing the facial expression of a query face image is in the order of $O(kWN_S)$+ $O(2\eta N_S D^2 W t_{all})$. Experiments were conducted on a computer system with Pentium IV 2.4GHz CPU and 512MB RAM. The average times required to compute the edge map, Gabor map and angle map of a face image are about 0.001 s, 0.10 s and 2.4×10^{-4} s, respectively. The average runtimes for feature matching using our SMOM-ESTM algorithm based on the AR database (3 expression models, 3 mean images as model images, and 183 images for testing) and the Yale database (5 expression models, 5 mean images as model images, and 75 images for testing) are 0.10 s and 0.17 s, respectively.

7.6 Conclusions

In this chapter, we have proposed a novel and accurate algorithm for human facial expression recognition. In our algorithm, a statistical model, namely Spatially Maximum Occurrence Model (SMOM), is proposed to model the different facial expressions, and the distance between a query input and an expression model is a measure of the precision of using the model to represent the expression in the query input. Another method, ESTM, is used to measure the similarity based on the shape and texture information using the shape-texture Hausdorff distance between the input image and the mean images of the training set. These two methods are

combined to form the SMOM-ESTM algorithm, which can achieve a good performance level for expression recognition.

In our algorithm, only the position of the eyes and the middle of the mouth are required for normalization and alignment. The expression models produced by SMOM contain most of the significant visual content in the training data, and are suitable for representing the expression patterns, which have large within-class variations. For ESTM, the shape information and texture information about an image are complementary and supplementary to each other, which can provide a more detailed and exact description of a facial expression. Furthermore, the elastic matching allows expression to vary to a certain extent. With our approach, the recognition rates based on the AR database and the Yale database are 94.5% and 94.7%, respectively.

Chapter 8. Human Face Indexing

In the previous chapter, we proposed different methods for face recognition, lighting modeling and facial expression recognition. All the methods are important for the development of a practical and reliable face recognition system. However, human faces are usually represented by a high dimensional feature vector. The computation required for face recognition will become prohibitively large if the size of the database is very large. In this chapter, we will propose an efficient indexing structure for searching for a human face in a large database.

8.1 Introduction

As more and more information is captured and stored in digital form, the requirement of digital libraries/databases has significantly increased, for example, video data, human face images, etc. In the meantime, computational complexity for indexing and retrieving the information will become prohibitively heavy when the size of the database concerned becomes very large. The range of applications for a digital library is wide, covering electronic commerce, security, human computer interaction, etc. Human face recognition is an example of such applications, and it usually involves a large database that can have a size of thousands. However, the computational time for retrieving a matched human face from a database will increase with the size of the face database. Hence, efficient indexing of human faces in a large database is an important issue in making the application practical. In a face recognition system, each human face in the database is pre-processed and

its feature for recognition is extracted. This feature vector is stored along with the corresponding face image. With a query input face image, the same type of feature is extracted and then compared to each of those in the database. The similarity between the query input and a face in the database is measured by the distance between their respective feature vectors. Therefore, the runtime required for the face recognition process can be reduced if the number of face images in the database to be considered is smaller.

Currently, there are many techniques for quick image retrieval or image indexing, such as the color-based approach [202, 203] and shape-contour retrieval [204-206]. However, these methods may not be applied for indexing face images in a database because each human face has a similar facial shape and color. In this chapter, we introduce a new efficient indexing algorithm for face recognition with a large database. This is a two-stage approach. In the first stage, a small set of faces in the database similar to the input is selected to a smaller database, namely a condensed database, with the computation to be required independent of the database size. Then, in the second stage, a more accurate but more computational method is applied to search for the required face in the condensed database.

8.2 Indexing for a Human Face Database

The purpose of indexing is to allow for the retrieving of required images from a database quickly. Usually, when the size of the feature vector is less than 20, many efficient indexing schemes based on tree structures can be used. However, for human face recognition, the dimension of the feature vector to represent a human face is much larger than 20. In our approach, eigenfaces are used

to form a vantage-object structure [207], which can efficiently select similar faces in a database. This can therefore reduce a large database problem to a small one, and can afford to use an accurate yet computational face recognition method in the second stage.

8.2.1 Eigenfaces as Vantage Object

As discussed in Section 2.1.3.1, PCA has been a popular technique for human face recognition. The eigenvectors obtained by PCA, which have the best representation of the original training faces, are called eigenfaces. This technique can also reduce the dimension of the input image to a dimension depending on the number of eigenfaces being used to represent the image. The input image is projected onto the eigenfaces to form a feature vector for its representation and face recognition.

Suppose that each face is normalized to a size of $N \times M$, and the face images are denoted as $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, $\mathbf{\Gamma}_{3,\ldots,}$ $\mathbf{\Gamma}_k$, where k is the number of face images stored in a database. Then, the average face, $\boldsymbol{\psi}$, and difference faces, $\mathbf{\Phi}_i$, are defined as follows: $\mathbf{\Phi}_i = \mathbf{\Gamma}_i - \boldsymbol{\psi}$, where

$$\boldsymbol{\psi} = \frac{1}{k}\sum_{i=1}^{k} \mathbf{\Gamma}_i \ . \tag{8.1}$$

Then, the covariance matrix Ω, which is of dimension $NM \times NM$, is computed as follows:

$$\Omega = \Phi \cdot \Phi^T , \tag{8.2}$$

where $\Phi = [\Phi_1 \ \Phi_{2\ldots}\Phi_k]$, a $NM \times k$ matrix. The eigenvectors and the corresponding eigenvalues of the covariance matrix can be computed as follows:

$$\mathbf{\Omega v}_i = \lambda_i \mathbf{v}_i. \quad (8.3)$$

The covariance matrix has up to P eigenvectors associated with non-zero eigenvalues, where P = minimum(NM, k). The eigenvectors $\mathbf{v}_i$ are sorted from high to low according to their associated eigenvalues λ_i. The set of eigenvectors V represents the principal components of the training face images, and the eigenvectors are denoted as follows:

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_P]. \quad (8.4)$$

A normalized difference face image, φ, is projected onto the eigenspace to form a feature vector, κ, of dimension P.

$$\boldsymbol{\kappa} = [\kappa_1, \kappa_2, \ldots, \kappa_P]^T = \mathbf{V}^T \cdot \boldsymbol{\varphi}. \quad (8.5)$$

The i^{th} value of κ is computed as the dot product of the face image, φ, and the i^{th} eigenvector, $\mathbf{v}_i$. This feature vector κ can then be used to represent the input face image for face recognition. In practice, the actual number of eigenvectors or eigenfaces to be used, p, is much smaller than P. In our indexing structure, the magnitude of each projection, κ_i, is used individually in ranking the face images to form p ranked lists. With different eigenfaces, the face images are ranked in different orders.

8.2.2 Formation of a Condensed Database

The computation required to search a query face image in a database is a function of the feature vector dimension and the number of subjects in the database. Therefore, the search process will become much faster if a small sub-set of the faces in the database can be selected efficiently to form a condensed database, which also includes the matched face. Therefore, in our indexing scheme, a condensed database is generated from a large database for a query face image.

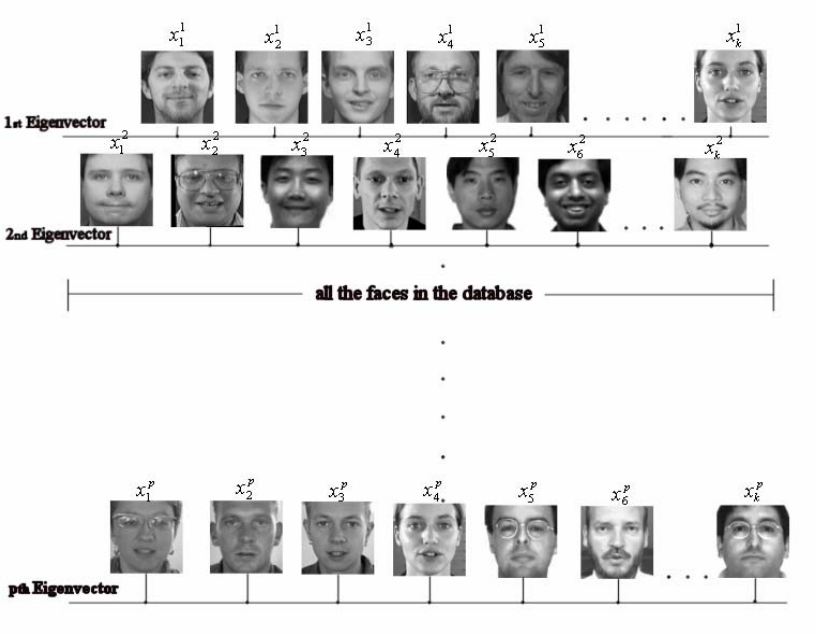


Figure 8-1 The structure of our indexing scheme. The faces in the database are ranked with respect to p eigenfaces to form p ranked lists.

Suppose that each face image in the database, $\mathbf{\Gamma}_i$, is normalized and subtracted by the average face, and projected onto the p eigenfaces with the corresponding largest eigenvalues. Therefore, the face image is decomposed into p projected values $[\kappa_{i,1},\kappa_{i,2}, \ldots,\kappa_{i,p}]$. For each of the eigenfaces, the face images are ranked in either an ascending or descending order. In the following, the discussion assumes that the faces are ranked in ascending order, and each face is ranked p times to form p ranked lists. Figure 8-1 illustrates the indexing scheme with p eigenfaces. The projected value onto the m^{th} eigenface with corresponding rank j is denoted as x_j^m. The rank of a face with respect to an eigenface depends on its similarity to the eigenface relative to other face images in the database. Similar faces should have similar ranks with respect to the eigenfaces. In our method, a query image is normalized and then ranked with respect to each of the eigenfaces.

Suppose that **Y** denotes the normalized difference face. Its projections onto each of the eigenfaces $\mathbf{v}_i$ are computed as follows:

$$y_i = \mathbf{v}_i^{\mathrm{T}} \cdot \mathbf{Y}, \qquad \text{where } i = 1, \ldots, p. \tag{8.6}$$

The input **Y** is then ranked in the p ranked lists. Similar face images in the database are then selected by considering its neighbors in each of the ranked list, as illustrated in Figure 8-2. Suppose that the input is ranked between position j and j+1 in the m^{th} ranked list. The one with its projected value, i.e. x_j^m or x_{j+1}^m, nearest to y_m is selected and then put into the condensed database B. Similar faces from the database are selected by considering the p ranked lists one by one and repetitively until the condensed database B containing the required number of distinct faces.

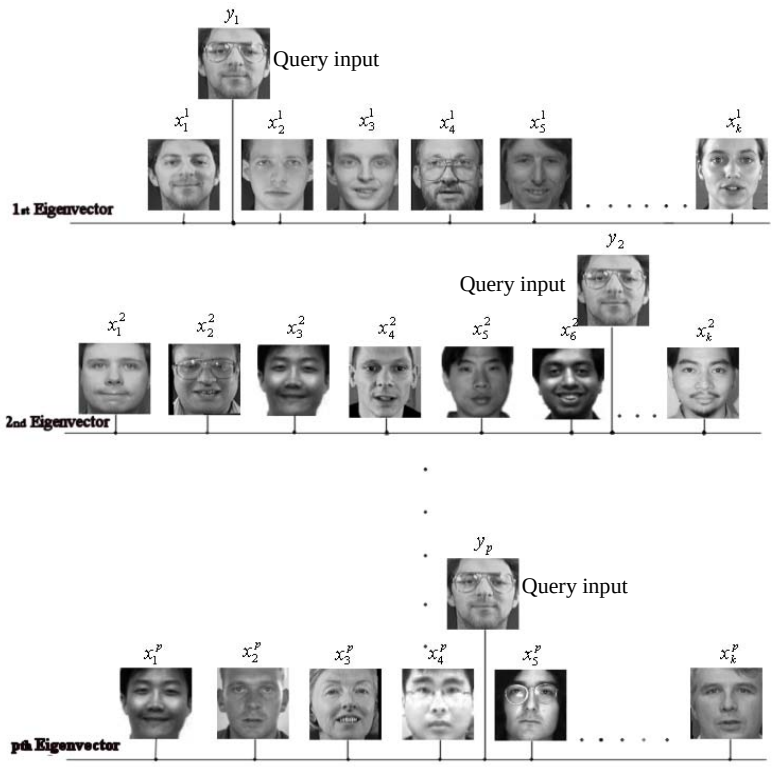


Figure 8-2 Ranking a query face image and selecting similar faces to the condensed database B.

The computation required for this scheme depends on the number of eigenfaces to be used. The smaller number of eigenfaces used, the less the amount of computation required. However, a larger condensed database *B* will then be needed in order to guarantee the inclusion of the matched face to the query. Furthermore, the number of eigenfaces to be used and the size of the condensed database *B* depend on the size of the large face database concerned.

8.3 Experimental Results

8.3.1 Indexing Using Eigenfaces

To investigate the performance of our proposed indexing scheme based on eigenfaces, a number of standard face databases and self-captured face images were used in the experiments. Those databases used include the ORL database, Yale face database, MIT face database [208], AR face database, BioID face database [209], UMIST face database [210], and Bern face database [211]. Including our self-captured ones, we form a database of 523 distinct subjects, with 752 different facial images for testing. The face images of each subject were captured at different times, under slightly different lighting conditions, and with slightly perspective variations.

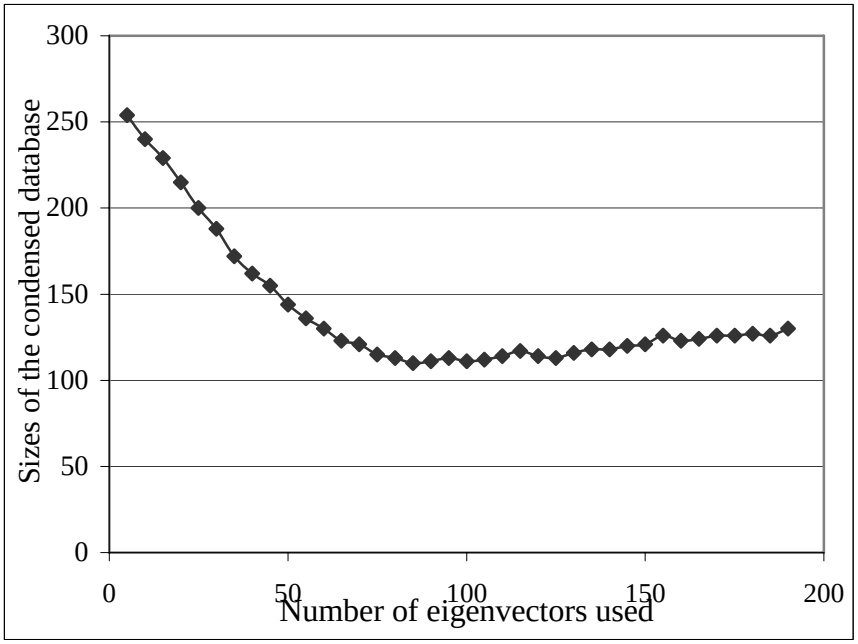
The experiment setup is that an upright frontal view of each of the 523 subjects with a suitable scale and normal facial expression was chosen to form a database consisting of 523 persons. Some of the remaining face images are selected to form a testing set of 752 faces. The eigenfaces based on the 523 face images in the database are generated, and our proposed indexing structure is formed with different numbers of eigenfaces. In our experiments, we will investigate the

number of eigenfaces to be used such that the size of the condensed database will be a minimum. With an optimal number of eigenfaces, the size of the condensed database, which guarantees the inclusion of the matched face, will be investigated. The objective is, with a certain sized large face database, to investigate the corresponding number of eigenfaces to be used and the size of the condensed database to achieve the best performances in terms of computation and recognition rate. In the following experiments, the effect of database size on the optimal number of eigenfaces to be used and the size of the condensed database will be studied, and two different database sizes, 330 and 523, will be considered.

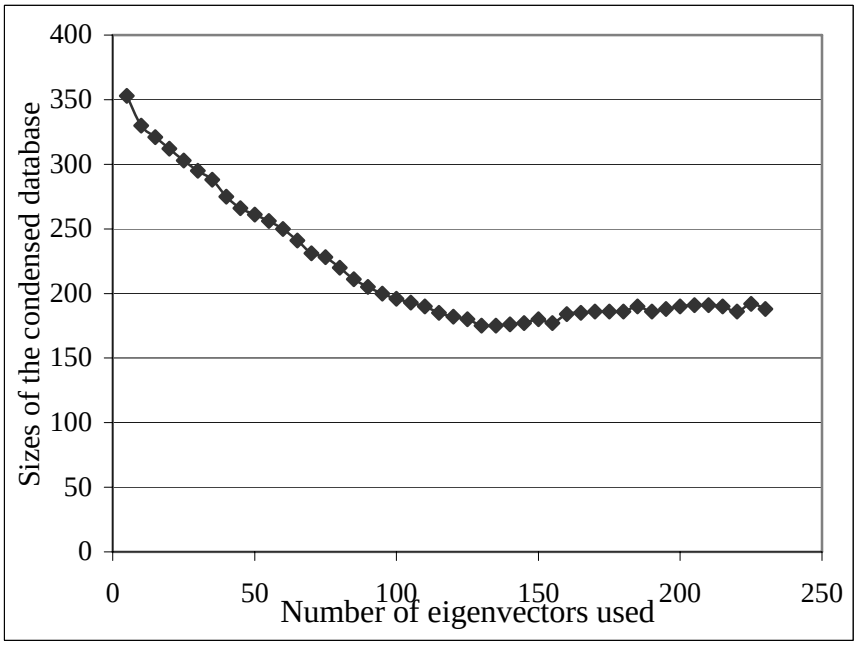
With a particular database size, different numbers of eigenfaces used will affect the required size of the condensed database. Figures 8-3(a) and 8-3(b) illustrate the number of eigenvectors used in our indexing scheme and the corresponding size of the condensed database which will include all the matching faces among the testing faces. From the experimental results, the required size of the condensed database decreases with an increase in the number of eigenfaces being used, until a certain number. These are 83 and 128, when the database sizes are 330 and 523, respectively. In order words, the optimal number of eigenfaces to be used is roughly 25% of the total number of different faces in a database.

When the size of a database grows, the corresponding size of the condensed database will also increase. From Figure 8-3, the size of the condensed database should be set at 35% of the size of the database. The runtime required by our indexing scheme to produce the condensed database depends on the number of eigenfaces used. Experiments show that the average runtimes for using 25% of

eigenfaces with a database size of 523 is less than 1 second. The experimental results were conducted on a Pentium 4 2.4GHz computer system.



(a)

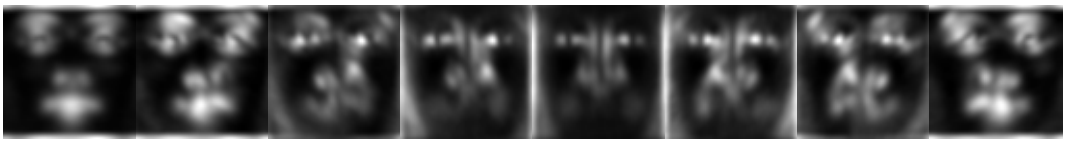


(b)

Figure 8-3 The required sizes of the condensed database with different numbers of eigenfaces used in our indexing scheme when the sizes of the database concerned are (a) 330 and (b) 523, respectively.

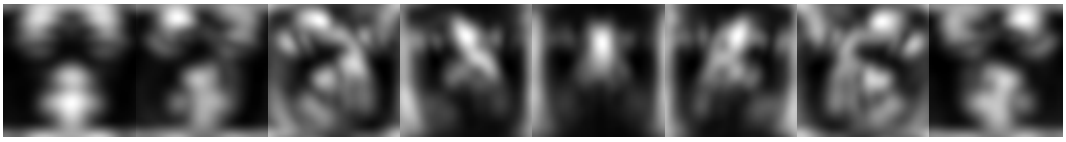
Indexing Using Gaborfaceshis part, Gaborfaces, which are produced by using the Gabor wavelets, are used instead of eigenfaces for human face indexing. The performance of this approach will also be evaluated and compared to the eigenface approach.

The Gabor representations of an input image are the convolution outputs of the image and a set of Gabor wavelet filters [2.19] with different center frequencies and orientations. Some Gabor representations are shown in Figure 2-1. In our experiments, we select the three center frequencies $\left(\pi/4, \sqrt{2}\pi/4, \pi/2\right)$ with a scale factor of $\sqrt{2}$ [54]. In addition, 8 orientations are used in our experiments. Therefore, 24 Gabor wavelet kernels are used to extract the features in an image. For each training image, 24 outputs are generated based on the different Gabor wavelet kernels. For each kernel, an average representation, which is called a Gaborface, is computed. Then, we use these Gaborfaces instead of the eigenfaces in section 8.3.1 to build a human face indexing scheme. As 3 scales and 8 orientations are used, we also compute 3 average images for the three different scales, and 8 average images for the eight different orientations. In totally, we have 24+3+8=35 Gaborfaces for human face indexing. In our experiments, 433 images were selected as training images, and the Gaborfaces generated are shown in Figure 8-4. These Gaborfaces are similar to the Gabor representations shown in Figure 2-1, because the former are the average images of the latter based on different identities. Compared to the eigenface approach, the Gaborfaces retain more local features, which exhibit the desirable characteristics of capturing salient visual properties such as spatial localization, orientation selectivity, and spatial frequency.



(a) Gaborfaces with different orientations and $\pi/4$ center frequency

(b) Gaborfaces with different orientations and $\sqrt{2}\pi/4$ center frequency



(c) Gaborfaces with different orientations and $\pi/2$ center frequency

(d) Gaborfaces with three different center frequencies

(e) Gaborfaces with eight different orientations

Figure 8-4 The magnitudes of the Gaborfaces with 3 scales and 8 orientations.

It is not necessary to use all these Gaborfaces in face indexing. Four different configurations of the Gaborfaces are used in our experiments. Configuration A selects the first 24 Gaborfaces which have different scales and orientations; Configuration B selects the 3 average Gaborfaces with different center frequencies; Configuration C selects the 8 average Gaborfaces with different orientations, and Configuration D considers all the total 35 Gaborfaces. In our database, 282 images are used for testing. The respective experimental results are shown in Figure 8-5.

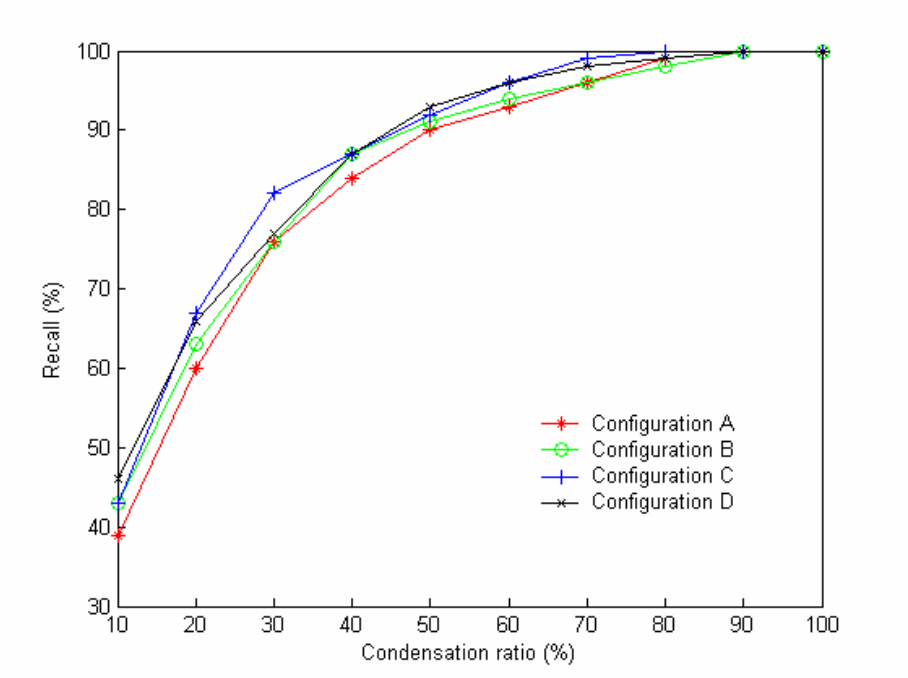


Figure 8-5 Human face indexing using Gaborfaces.

In Figure 8-5, the condensation ratio measures the condensation efficiency of the testing database; recall indicates the proportion of desired images that are returned among the condensed database. Although different numbers of Gaborfaces are used, the results based on these four configurations are very similar. The use of Configuration C achieves the best performance. When the size of the condensed database is set at 35% of the size of the database, the recall of the testing images is about 85%. Compared to the performance of the human face indexing scheme using eigenfaces, this result is not satisfactory. This can be explained by the fact that the Gaborfaces mainly consider the local texture of the image, and consequently are sensitive to scaling, translation, and the variations of expressions, etc. Therefore, for some query images, no matter which Gaborface is used, more neighbors in the ranked list should be considered. [109] used the line edge map (LEM) to prefilter a face database, and for the AR face database, when the size of

the condensed database is set at 49.69% of the size of the database (i.e. 50.31% of images are filtered out), the recall rate of the testing images is about 88.39%. This performance is not only lower than our method based on the Gaborfaces (recall rate of about 92% for 50% condensation ratio), but also much lower than the result based on the eigenfaces (recall rate of 100% for 35% condensation ratio).

8.4 Conclusions

In this chapter, a new efficient indexing scheme, which produces a condensed database from a large face database, is proposed. Our approach is based on eigenfaces and the projections of a face image onto each of the eigenfaces are used for its ranking. The computational complexity of this scheme is proportional to the number of eigenfaces used. Experimental results show that the optimal number of eigenfaces to be used and the size of the condensed database are about 25% and 35%, respectively, of the size of the database. This allows us to consider a small condensed database instead of the original large face database when performing face recognition. We have also proposed another human face indexing scheme using Gaborfaces, and discussed the reason why its performance is lower than that of using eigenfaces.

Chapter 9. Conclusions and Future Work

9.1 Conclusion on our current work

In this thesis, we first describe the principles of face image analysis techniques. Our research focuses on three areas: face recognition, face recognition under varying illuminations, and facial expression recognition. We make a brief review on some well-known face recognition techniques; and also review the recent developments of the methods for face recognition under varying illuminations and the methods for facial expression recognition.

We propose two methods for face recognition under various conditions: ESTM and DKPCA. ESTM, elastic shape-texture matching method, is devised in Chapter 3. In this approach, not only the shape information but also the texture information is used for comparing two faces without establishing any precise pixel-wise correspondence. Because the elastic matching is carried out within the neighborhood of each edge pixel concerned, which is robust to small, local distortions of the feature points, such as facial expression variations, this method is robust to small shape variations. The edge map, Gabor representations and the direction of image gradient can all alleviate the effect of illumination to a certain extent. However, when violent illumination variations exist, the edge map is not reliable, and the performance of ESTM will degrade. DKPCA, Doubly nonlinear mapping kernel Principal Component Analysis, is a Gabor-based method and is proposed in Chapter 4. In our approach, the Gabor wavelets are used to extract

facial features, then a doubly nonlinear mapping kernel PCA is proposed to perform feature transformation and face recognition. The proposed nonlinear mapping not only considers the statistical property of the input features, but also adopts an eigenmask to emphasize those important facial feature points. Therefore, after this mapping, the transformed features have a higher discriminating power, and the relative importance of the features adapts to the spatial importance of the face images. This new nonlinear mapping is combined with the conventional kernel PCA for face recognition.

The experiments using ESTM and DKPCA are performed, and the corresponding performances are compared with other methods in Chapter 3 and Chapter 4, respectively. From Table 3-8 and Table 4-6, we can see that DKPCA always outperforms ESTM; this is because the former can encode higher order statistics and the features are recoded according to their statistical property and shape importance. However, if there are large expression variations exist, e.g. see the ORL database in Table 3-6 and Table 4-4, ESTM can perform better. This is because the elastic matching is more suitable for comparison under local shape variations. Here we should point out that in order to compare the performances of different methods with the same conditions, all methods shown in Table 3-8 adopt the Euclidean distance measure for computing the similarity, while the methods in Table 4-6 use the Mahalanobis distance measure. It is difficult to determine which distance measure is better for a method. Although Liu [50] and Yambor et al. [185] argued that a PCA classifier will perform better when the Mahalanobis distance is used, the result in Table 3-8 and Table 4-6 show that, if there are large illumination

variations exist, e.g. the YaleB database, the Euclidean distance measure performs better.

Besides the methods for face recognition under various conditions, we also investigate the techniques for face recognition under varying illuminations, which can be considered a sub-problem of the former. We propose two methods, which are describes in Chapter 5 and Chapter 6, respectively. These two methods are model-based methods. According the illumination model and human face model we used, the effect of uneven illumination can be modeled by a sequence of multiplicative factors and additive factors, which are only determined by the illumination model concerned and the shape of a human face. Firstly, we propose a method, which can compensate for the uneven illuminations on human faces and reconstruct face images in normal lighting conditions, in Chapter 5. In order to eliminate the influence of shape about different faces, a 2D face shape model is used to obtain a shape-free texture image. For an identified illumination category, the effects of the uneven lighting, i.e. the multiplicative factor and the additive factor, can be computed using a set of training images, and are used for reconstructing an image that will visually be under normal illumination. Then these images can be used for face recognition. This method can produce a great visual improvement and lighting smoothness. However, it is difficult determine the feature points when the input image is under varying lightings. Therefore, we propose another illumination compensation method in Chapter 6, which is much simpler and efficient. Instead of computing the multiplicative factors and the additive factors, we aim to reduce or even remove the effect of these factors. In our method, a local normalization (LN)

technique is applied to an image, which can effectively and efficiently eliminates the effect of uneven illuminations while keeping the local statistical properties of the processed image the same as in the corresponding image under normal lighting condition. After processing, the image under varying illumination will have similar pixel values to the corresponding image that is under normal lighting condition. Then, the processed images are used for face recognition.

The results shown in Table 5-3 and Table 6-2 are similar. It seems that the method using LN technique is more efficient and effective than the method based on a 2D shape model. However, the latter method still has its advantages. Firstly, due to the shape normalization, it is robust to local distortion of a human face, such as the expression variations. Secondly, it can construct a visually nature human face, which is under normal illumination, therefore the results can be used for image reconstruction. Finally, in the Chapter 5, due to limitation of shape morphing, we only consider the central facial areas (see Figure 5-5, 5-6 and 5-7), which do not include the cheek and chin areas, and are more difficult for face recognition. In Chapter 3 and Chapter 4, we also evaluate the ESTM and DKPCA for face recognition under varying illuminations, and the results are shown in Table 3-5 and Table 4-3, respectively. The results are also similar to those in Table 5-3 and Table 6-2. We should notice that the methods in Chapter 3 and Chapter 4 are Gabor-based methods, while the methods in Chapter 5 and Chapter 6 directly perform the PCA on the intensity images. Table 6-4 shows the results if we adopt the Gabor wavelets to extract features based on the images processed by LN technique, the recognition rate is 98.7% for the combined database.

We also present an efficient method for human facial expression recognition in Chapter 7. We first propose a representation model for facial expressions, namely spatially maximum occurrence model (SMOM), which is based on the statistical characteristics of training facial images and has a powerful representation capability. The ESTM algorithm is then used to measure the similarity between images for facial expression recognition. By combining SMOM and ESTM, the algorithm is called SMOM-ESTM and can achieve a higher recognition performance level.

In order to reduce the computational time when perform face recognition based on a large-scale database, it is necessary to prefilter a smaller face database, which includes the target. In Chapter 8, an efficient indexing structure for searching a human face in a large database is proposed. In our method, a set of eigenfaces is computed based on the faces in the database. Each face in the database is then ranked according to its projection onto each of the eigenfaces. A query input will be ranked similarly, and the corresponding nearest faces in the ranked position with respect to each of the eigenfaces are selected from the database. These selected faces will then form a small database, namely a condensed database, for face recognition, instead of considering the original large database.

9.2 Future Work

9.2.1 Face Recognition Using Morphable Models

For images with non-frontal presentation, the 2002 Face Recognition Vendor Test (FRVT 2002) [212] examined the use of morphable models — a technique that takes a facial image from any angle and projects what the subject might look like

facing forward. It has been shown that there is a dramatic improvement in performance using the morphable models. One of the top three systems increased its performance from 26 percent on non-processed, non-frontal images to 84 percent on morphed images. In fact, in our illumination compensation algorithm proposed in Chapter 5, a morphable face model is also used. In order to build the correspondence between the query image and the reference image, similar to [170], a set of feature points is determined manually. Then, the displacements of these key points are computed. With this correspondence, a triangle-based cubic interpolation method is used to build a displacement field on the whole face. In order to build this correspondence automatically, the edge detection method and our facial expression analysis algorithm will be used. Therefore, we can automatically detect the locations of feature points, and use the morphable face model for face analysis.

Zhang et al. [213] creates a 3-D face structure from multiple image views of a human face taken in prior unknown poses by appropriately morphing a generic 3-D face. In fact, our morphable face model can also be extended from 2-D to 3-D. Nevertheless, the key issue to be investigated is how to build the correspondence between different images with varying poses.

9.2.2 Face Recognition Under Various Illuminations and with Different Expressions

As stated by Adini *et al*. [67], “The variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity”, the varying illuminations, expressions and perspectives are great challenges for automatic face recognition techniques. In

Chapters 5 and 6, we propose two methods to handle the illumination problem, and in Chapter 7 we describe a method to analyze facial expressions. When we consider the effect of varying illumination, we do not consider the expression variations, and *vice versa*. For a real face recognition application, these two factors may appear at the same time and affect each other. Therefore, we should propose a method which is robust to both the illumination and the expressions. A possible two-step procedure is to first perform the illumination normalization using the LN technique, and then to analyze the facial expression and perform the shape normalization, where a feature point detection technique is required.

For variations caused by perspectives or poses, it is difficult to recognize an identity based on a single frontal training image, and the multi-view techniques are required. In this case, morphable models can improve the performance as stated in Section 9.2.1. Therefore, how to combine these techniques to produce an automatic face analysis system which can analyze the illumination, facial expression and pose is an interesting and challenging research topic.

9.2.3 Other Applications of Human Face Analysis

Our human face analysis techniques, such as the recognition and representation of facial expressions, compensation for illumination conditions, use of texture and shape information for the morphable models, etc., can be applied for many purposes, e.g. face image compression, face image denoising, face image enhancement or super-resolution, and face image reconstruction. In these applications, we should combine our methods with other techniques according to the characteristics of the respective applications.

References

[1] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Survey*, vol. 35, no. 4, pp. 399-458, 2003.

[2] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," *IEEE Proceedings*, vol. 83, no. 5, pp. 705-741, 1995.

[3] S.G. Kong, J. Heo, B.R. Abidi, J. Paik and M.A. Abidi, "Recent Advances in Visual and Infrared Face Recognition - A Review," *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 103-135, 2005.

[4] J.N.K. Liu, M. Wang, and B. Feng, "IBotGuard: An Internet-based Intelligent Robot Security System Using Invariant Face Recognition Against Intruder," *IEEE Transactions Systems, Man and Cybernetics, Part C*, vol. 35, no. 1, pp. 97-105, 2005.

[5] A. Samal and P.A. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65-77, 1992.

[6] A.K. Jain, A. Ross, S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Transactions Circuits and Systems for Video Technology*, vol. 14, pp. 4-28, 2004.

[7] K. Delac, M. Grgic, "A Survey of Biometric Recognition Methods," *Proceedings of the 46th International Symposium Electronics in Marine, ELMAR-2004*, pp. 184-193, 2004.

[8] S.V. Ioannou, A.T. Raouzaiou, V.A. Tzouvaras, T.P. Mailis, K.C. Karpouzis and S.D. Kollias, "Emotion Recognition Through Facial Expression Analysis Based on a Neurofuzzy Network," *Neural Networks*, in press, available online 15 June 2005.

[9] V. Vezhnevets, "Face and Facial Feature Tracking for Natural Human-Computer Interface," *Proceedings of Graphicon 2002*, Nizhny Novgorod, Russia, pp. 86-90, 2002.

[10] M. Lockie, "The Biometric Industry Report - Forecasts and Analysis to 2006," In: Elsevier Advanced Technology, 2002.

[11] M.H. Yang, D.J. Kriegman, N. Ahuja, "Detecting Faces in Images: A Survey," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp.34-58, 2002.

[12] G. Yang and T. S. Huang, "Human Face Detection in Complex Background," *Pattern Recognition*, vol. 27, no. 1, pp.53-63, 1994.

[13] S. Mckenna, S. Gong, and Y. Raja, "Modeling Facial Colour and Identity with Gaussian Mixtures," *Pattern Recognition*, vol. 31, no. 12, pp. 1883-1892, 1998.

[14] K.C. Yow and R. Cipolla, "Feature-Based Human Face Detection," *Image and Vision Computing*, vol. 15, no. 9, pp. 713-735, 1997.

[15] H. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23-38, 1998.

[16] A.N. Rajagopalan, R. Chellappa, N.T. Koterba, "Background Learning for Robust Face Recognition with PCA in the Presence of Clutter," *IEEE Transactions Image Processing*, vol. 14, no. 6, pp. 832-843, 2005.

[17] H. Cevikalp, M. Neamtu, M. Wilkes, A. Barkana, "Discriminative Common Vectors for Face Recognition," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 4-13, 2005.

[18] F. Marqués and V. Vilaplana, "Face Segmentation and Tracking Based on Connected Operators and Partition Projection," *Pattern Recognition*, vol. 35, no. 3, pp. 601-614, 2002.

[19] Y. Fan, M. Paindavoine, "Implementation of an RBF Neural Network on Embedded Systems: Real-Time Face Tracking and Identity Verification," *IEEE Transactions Neural Networks*, vol. 14, no. 5, pp. 1162-1175, 2003.

[20] B. Fasel and J. Luettin, "Automatic Facial Expression Analysis: A Survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259-275, 2003.

[21] M. Pantie and L.J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424-1445, 2000.

[22] I.A. Essa and A.P. Pentland, "Coding, Analysis, Interpretation, and Recognition of Facial Expressions," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 757-763, 1997.

[23] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski, "Classifying Facial Actions," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974-989, 1999.

[24] Y.I. Tian, T. Kanade, and J.F. Cohn, "Recognizing Action Units for Facial Expression Analysis," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97-115, 2001.

[25] P. Ekman and W.V. Friesen, “Facial Action Coding System: A Technique for the Measurement of Facial Movement,” Consulting Psychologists Press, Palo Alto, 1978.

[26] B. Moghaddam, M.H. Yang, “Learning Gender with Support Faces,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 707-711, 2002.

[27] S. Gutta, H. Wechsler, “Gender Classification of Human Faces Using Hybrid Classifier Systems,” *Proceedings of the International Conference on Neural Networks*, vol. 3, pp. 1353-1358, 1997.

[28] D. Valentin, H. Abdi, B. Edelman and A.J. O'Toole, “Principal Component and Neural Network Analyses of Face Images: What Can Be Generalized in Gender Classification?” *Journal of Mathematical Psychology*, vol. 41, no. 4, pp. 398-413, 1997.

[29] A. Lanitis, C.J. Taylor, T.F. Cootes, “Toward Automatic Simulation of Aging Effects on Face Images,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442-455, 2002.

[30] R. Iga, K. Izumi, H. Hayashi, G. Fukano, T. Ohtani, “A Gender and Age Estimation System from Face Images,” *SICE 2003 Annual Conference*, vol. 1, pp. 756-761, 2003.

[31] C. Choi, “Age Change for Predicting Future Faces,” *Proceedings of the IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 99*, vol. 3 pp. 1603-1608, 1999.

[32] J. Zhao, Y. Su, D. Wang and S. Luo, “Illumination Ratio Image: Synthesizing and Recognition with Varying Illuminations,” *Pattern Recognition Letters*, vol. 24, no. 15, pp. 2703-2710, 2003.

[33] L. Tang, T.S. Huang, “Face Recognition Using Synthesized Intermediate Views,” *Proceedings of the38th Midwest Symposium on Circuits and Systems*, vol. 2, pp. 1066-1069, 1995.

[34] A. Tsukamoto, C.W. Lee, S. Tsuji, “Pose Estimation of Human Face Using Synthesized Model Images,” *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 93-97, 1994.

[35] M. Chan, C.Y. Chen, G. Barton, P. Delmas, G. Gimel'farb, P. Leclercq, T. Fischer, “Evaluation of 3D Face Analysis and Synthesis Techniques,” *Proceedings of IEEE International Conference on Multimedia and Expo*, vol. 2, pp. 1203-1206, 2004.

[36] D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang and W. Gao, “Efficient 3D Reconstruction for Face Recognition,” *Pattern Recognition*, vol. 38, no. 6, pp. 787-798, 2005.

[37] M.W. Lee and S. Ranganath, “Pose-Invariant Face Recognition Using a 3D Deformable Model,” *Pattern Recognition*, vol. 36, no. 8, pp. 1835-1846, 2003.

[38] Y.Z. and C. Kambhamettu, “3D Head Tracking Under Partial Occlusion,” *Pattern Recognition*, vol. 35, no. 7, pp. 1545-1557, 2002.

[39] L. Sirovich and M. Kirby, “Low-Dimensional Procedure for the Characterization of Human Faces,” *Journal of the Optical Society of America A*, vol.4, pp. 519-524, 1987.

[40] M. Kirby and L. Sirovich, “Application of the KL Procedure for the Characterization of Human Faces,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103-108, 1990.

[41] M. Turk and A. Pentland, “Eigenfaces for Recognition,” *Journal of Cognitive Neuroscience*, vol. 3, pp. 71-86, 1991.

[42] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.

[43] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, “Face Recognition by Independent Component Analysis,” *IEEE Transactions Neural Networks*, vol. 13, no. 6, pp. 1450-1464, 2002.

[44] C. Liu and H. Wechsler, “Independent Component Analysis of Gabor Features for Face Recognition,” *IEEE Transactions Neural Networks*, vol. 14, no. 4, pp. 919-928, 2003.

[45] O. Déniz, M. Castrillón, M. Hernández, “Face Recognition Using Independent Component Analysis and Support Vector Machines,” *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2153-2157, 2003.

[46] B.A. Draper, K. Baek, M.S. Bartlett and J.R. Beveridge, “Recognizing Faces with PCA and ICA,” *Computer Vision and Image Understanding*, vol. 91, no. 1-2, pp. 115-137, 2003.

[47] B. Scholkopf and A. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond. MIT Press, 2002.

[48] B. Scholkopf, A. Smola, and K. R. Muller, “Nonlinear Component Analysis as a Kernel Eigenvalue Problem,” *Neural Computation*, vol. 10, no. 5, pp. 1299-1319, 1998.

[49] B. Moghaddam, “Principal Manifolds and Probabilistic Subspaces for Visual Recognition,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 780-788, 2002.

[50] C. Liu, “Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 572-581, 2004.

[51] D.P. Huttenlocher, G.A. Klanderman and W.J. Rucklidge, “Comparing Images Using the Hausdorff Distance,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850-863, 1993.

[52] M.P. Dubuisson and A.K. Jain, “Modified Hausdorff Distance for Object Matching,” *Proceedings of the 12th International Conference on Pattern Recognition*, Jerusalem, Israel, pp. 566-568, 1994.

[53] C.K. Chui, An Introduction to Wavelets, Boston: Academic Press, 1992.

[54] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C.V.D. Malsburg, R.P. Wurtz, and W. Konen, “Distortion Invariant Object Recognition in the Dynamic Link Architecture,” *IEEE Transactions Computers*, vol. 42, no. 3, pp. 300-311, 1993.

[55] K.W. Wong, K.M. Lam and W.C. Siu, “A Robust Scheme for Live Detection of Human Face in Color Images,” *Signal Processing*: *Image Communication*, vol. 18, no. 2, pp. 103-114, 2003.

[56] K.H. Lin, K.M. Lam and W.C. Siu, “Spatially Eigen-Weighted Hausdorff Distances for Human Face Recognition,” *Pattern Recognition*, vol. 36, no. 8, pp 1827-1834, 2003.

[57] A. Rajagopalan, K. Kumar, J. Karlekar, R. Manivasakan, M. Patil, U, Desai, P. Poonacha, and S. Chaudhuri, “Finding Faces in Photographs,” *Proceedings of the Sixth IEEE International Conference Computer Vision*, pp. 640-645, 1998.

[58] E. Osuna, R. Freund, and F. Girosi, “Training Support Vector Machines: An Application to Face Detection,” *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*, pp. 130-136, 1997.

[59] I. Craw, D. Tock, and A. Bennett, “Finding Face Features,” *Proceedings of the Second European Conference Computer Vision*, pp. 92-96, 1992.

[60] R. Kjeldsen and J. Kender, “Finding Skin in Color Images,” *Proceedings of the Second International Conference Automatic Face and Gesture Recognition*, pp.312-317, 1996.

[61] A. Kurosawa, N. Okamoto, "Automatic Extraction Method of Facial Regions for Occlusion Faces in Moving Pictures," *Canadian Conference Electrical and Computer Engineering*, vol. 1, pp. 329-334, 2001.

[62] P.J. Phillips, "Matching Pursuit Filters Applied to Face Identification," *IEEE Transactions Image Processing*, vol. 7, no. 8, pp. 1150-1164, 1998.

[63] J.G. Wang, E.T. Lim, R. Venkateswarlu, E. Sung, "Stereo Head/Face Tracking and Pose Estimation," *Proceedings of Seventh IEEE International Conference Automation, Robotics and Vision*, vol. 3, pp. 1609 -1614, 2002.

[64] X. Cui, Y. Li, R. Dai, "Face Pose Analysis and Estimation," *Proceedings of International Conference Control and Automation*, pp. 182-182, 2002.

[65] A. Tsukamoto, C.W. Lee, S. Tsuji, "Detection and Pose Estimation of Human Face with Synthesized Image Models," *Proceedings of International Conference Pattern Recognition*, vol. 1, pp. 754-757, 1994.

[66] P.N. Belhumeur and D.J. Kriegman, "What is the Set of Images of an Object under All Possible Lighting Conditions?" *Proceedings of International Conference Computer Vision and Pattern Recognition*, pp. 270-277, 1996.

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

[68] X. Xie and K.M. Lam, "Face Recognition under Varying Illumination Based on a 2D Face Shape Model," *Pattern Recognition*, vol. 38, no. 2, pp. 221-230, 2005.

[69] I.S. Bruner, AND R. Tagiuri, "The Perception of People," In *Handbook of Social Psychology*, vol. 2, G. Lindzey, Ed., Addison-Wesley, Reading, MA, pp. 634-654, 1954.

[70] W.W. Bledsoe, "The Model Method in Facial Recognition," Panoramic Research Inc., Tech. Rep. PRI:15, Palo Alto, CA, 1964.

[71] F. Galton, "Personal identification and description," *Nature*, pp. 173-188, June 21, 1888.

[72] M.D. Kelly, "Visual Identification of People by Computer," Tech. Rep. AI-130, Stanford AI Proj., Stanford, CA, 1970.

[73] T. Kanade, *Computer Recognition of Human Faces.* Birkhauser, Basel and Stuttgart, 1977.

[74] H.D. Ellis, "Introduction to Aspects of Face Processing: Ten Questions in Need of Answers," In *Aspects of Face Processing*, H. Ellis, M. Jeeves, F.

Newcombe, and A. Young, Eds. Nijhoff, Dordrecht, The Netherlands, pp. 3–13, 1986.

[75] I. Biederman and P. Kalocsai, "Neural and Psychophysical Analysis of Object and Face Recognition," In *Face Recognition: From Theory to Applications*, H.Wechsler, P.J. Phillips, V. Bruce, F.F. Soulie, and T.S. Huang, Eds. Springer-Verlag, Berlin, Germany, pp. 3–25, 1998.

[76] A. Pentland, B. Moghaddam, T. Stamer, and M. Turk, "View-based and modular eigenspaces for face recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 84-91, 1994.

[77] L. Wiskott, J.M. Fellous, N. Krüger, and C. Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-779, 1997.

[78] J. Yang; D. Zhang, A.F. Frangi, J. Yang, "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131-137, 2004.

[79] X. He, S. Yan, Y. Hu, P. Niyogi and H.J. Zhang, "Face Recognition Using Laplacianfaces," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328-340, 2005.

[80] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller, "Fisher Discriminant Analysis with Kernels," *Proc. IEEE International Workshop Neural Networks for Signal Processing IX*, pp. 41-48, Aug. 1999.

[81] G. Baudat and F. Anouar, "Generalized Discriminant Analysis Using a Kernel Approach," *Neural Computation*, vol. 12, no. 10, pp. 2385-2404, 2000.

[82] J. Lu, K.N. Plataniotis and A.N. Venetsanopoulos, "Face Recognition Using Kernel Direct Discriminant Analysis Algorithms," *IEEE Trans. Neural Networks*, vol. 14, no. 1, pp. 117-126, 2003.

[83] Q. Liu, H. Lu, and S. Ma, "Improving Kernel Fisher Discriminant Analysis for Face Recognition," *IEEE Transactions Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 42-49, 2004.

[84] J. Yang, A.F. Frangi, J. Yang, D. Zhang, and Z. Jin, "KPCA Plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 230-244, 2005.

[85] G. Guo, S.Z. Li, K. Chan, “Face Recognition by Support Vector Machines,” *Proceedings of the Fourth IEEE International Conference Automatic Face and Gesture Recognition*, pp. 196 – 201, 2000.

[86] B. Heisele, P. Ho, T. Poggio, “Face Recognition with Support Vector Machines: Global Versus Component-Based Approach,” *Proceedings of the eighth IEEE International Conference Computer Vision*, vol. 2, pp. 688-694, 2001.

[87] M. Safari, M.T. Harandi, B.N. Araabi, “A SVM-Based Method for Face Recognition Using a Wavelet PCA Representation of Faces,” *2004 International Conference on Image Processing*, vol. 2, pp. 853-856, 2004.

[88] V. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.

[89] J.C. Christorpher and Bugres, “A Tutorial on Support Vector Machines for Pattern Recognition,” *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, 1998.

[90] T.J. Stonham, Practical Face Recognition and Verification with WISARD, in: H.D. Ellis, M.A. Jeeves, F. Newcombe, A. Young (Eds.), Aspects of Face Processing, Kluwer, Boston, MA, pp. 426–441, 1986.

[91] S.H. Lin, S.Y. Kung, L.J. Lin, “Face Recognition/Detection by Probabilistic Decision-Based Neural Network,” *IEEE Transactions Neural Networks*, vol. 8, no. 1, pp. 114 – 132, 1997.

[92] S.Y. Kung, J.S. Taur, “Decision-Based Neural Networks with Signal/Image Classification Applications,” *IEEE Transactions Neural Networks*, vol. 6, no. 1, pp. 170 – 181, 1995.

[93] M.J. Er, S. Wu, J. Lu, H.L. Toh, “Face Recognition with Radial Basis Function (RBF) Neural Networks,” *IEEE Transactions Neural Networks*, vol. 13, no. 3, pp. 697-710, 2002.

[94] B. Duc, S. Fischer, and J. Bigun, “Face Authentication with Gabor Information on Deformable Graphs,” *IEEE Transactions Image Processing*, vol. 8, no. 4, pp. 504 - 516, 1999.

[95] C.L. Kotropoulos, A. Tefas, I. Pitas, “Frontal Face Authentication Using Morphological Elastic Graph Matching,” *IEEE Transactions Image Processing*, vol. 9, no. 4, pp. 555 - 560, 2000.

[96] J. Zhang, Y. Yan, and M. Lades, “Face Recognition: Eigenface, Elastic Matching, and Neural Nets,” *IEEE Proceedings*, vol. 85, no. 9, pp. 1423-1435, 1997.

[97] F. Samaria and F. Fallside, “Face Identification and Feature Extraction Using Hidden Markov Models,” *Image Processing: Theory and Application*, G. Vernazza, ed., Elsevier, 1993.

[98] F. Samaria and S. Young, “HMM-Based Architecture for Face Identification,” *Image and Vision Computing*, vol. 12, no. 8, pp. 537-543, 1994.

[99] A.V. Nefian, M.H. Hayes III, “Hidden Markov Models for Face Recognition,” *Proceedings of the IEEE International Conference Acoustics, Speech, and Signal Processing*, vol. 5, pp. 2721-2724, 1998.

[100] A.V. Nefian, M.H. Hayes III, “An Embedded HMM for Face Detection and Recognition,” *Proceedings of the IEEE International Conference Acoustics, Speech, and Signal Processing*, vol. 6, pp. 3553-3556, 1999.

[101] A.V. Nefian, “Embedded Bayesian Networks for Face Recognition,” *Proceedings of the IEEE International Conference Multimedia and Expo*, vol. 2, pp. 133-136, 2002.

[102] H. Othman, T. Aboulnasr, “A Separable Low Complexity 2D HMM with Aplication to Face Recognition,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1229 – 1238, 2003.

[103] X. Liu, T. Chen, “Video-Based Face Recognition Using Adaptive Hidden Markov Models,” *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*, vol. 1, pp. 340-345, 2003.

[104] R. Bruneli and T. Poggio, “Face Recognition: Features versus Templates,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1042-1052, 1993.

[105] B.S. Manjunath, R. Chellappa, and C. von der Malsburg, “A Feature Based Approach to Face Recognition,” *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*, pp. 373-378, 1992.

[106] S. Tamura, H. Kawa, H. Mitsumoto, “Male/Female Identification From 8×6 Very Low Resolution Face Images by Neural Network,” Pattern Recognition, vol. 29, no. 2, pp. 331-335, 1996.

[107] R.J. Baron, “Mechanism of Human Facial Recognition,” *International Journal of Man-Machine Studies*, vol. 15, pp. 137-178, 1981.

[108] M. Bichsel, “Strategies of Robust Object Recognition for Identification of Human Faces,” PhD thesis, Eidgenossischen Technischen Hochschule, Zurich, 1991.

[109] Y. Gao, M.K.H. Leung, "Face Recognition Using Line Edge Map," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 764-779, 2002.

[110] M.K.H. Leung and Y.H. Yang, "Dynamic Two-Strip Algorithm in Curve Fitting," *Pattern Recognition*, vol. 23, pp. 69-79, 1990.

[111] X. Xie and K.M. Lam, "Efficient Human Face Recognition Based on Shape and Texture", *Proceedings of the Seventh International Conference on Signal Processing, ICSP 2004*, vol. 2, pp. 938-941, 2004.

[112] C. Nastar and N. Ayach, "Frequency-Based Nonrigid Motion Analysis," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 18, pp. 1067-1079, 1996.

[113] A.J. Bell, T.J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Computation*, vol. 7, pp. 1129-1159, 1995.

[114] A. Hyvärinen, "The Fixed-Point Algorithm and Maximum Likelihood Estimation for Independent Component Analysis," *Neural Processing Letters*, vol. 10, no. 1, pp. 1-5, 1999.

[115] M.S. Bartlett, H.M. Lades, T.J. Sejnowski, "Independent Component Representations for Face Recognition," *SPIE Conference on Human Vision and Electronic Imaging* III, vol. 3299, pp. 528-539, 1998.

[116] M.S. Bartlett, T.J. Sejnowski, Viewpoint Invariant Face Recognition Using Independent Component Analysis and Attractor Networks, in: M. Mozer, M. Jordan, T. Petsche (Eds.), Neural Information Processing Systems – Natural and Synthetic, vol. 9, MIT Press, Cambridge, MA, 1997, pp. 817–823.

[117] C. Liu and H. Wechsler. "Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition," *Proceedings of the second International Conference on Audio and Video based Biometric Person Authentication*, Washington D. C., 1999.

[118] P.C. Yuen, J.H. Lai, "Independent Component Analysis of Face Images," *IEEE Workshop on Biologically Motivated Computer Vision*, Seoul, 2000.

[119] K. Baek, B.A. Draper, J.R. Beveridge, K. She, "PCA Vs ICA: A Comparison on the FERET Data Set," *Joint Conference on Information Sciences*, Durham, NC, 2002.

[120] B. Moghaddam, "Principal Manifolds and Bayesian Subspaces for Visual Recognition," *Proceedings of the Seventh IEEE International Conference Computer Vision*, vol. 2, pp. 1131-1136, 1999.

[121] W.S. Chen, P.C. Yuen, J. Huang, D.Q. Dai, “Kernel Machine-Based One-Parameter Regularized Fisher Discriminant Method for Face Recognition,” *IEEE Transactions on Systems, Man, and Cybernetics-Part B*: *Cybernetics*, Accepted for future publication, vol. PP, no. 99, pp. 659-669, 2005.

[122] Y. Zhu, E. Sung, “Margin-maximization discriminant analysis for face recognition,” *Proceedings of IEEE International Conference on Image Processing*, vol. 1, pp. 609-612, 2004.

[123] H. Yu and J. Yang, “A Direct LDA Algorithm for High-Dimensional Data — With Application to Face Recognition,” *Pattern Recognition*, vol. 34, no. 10, pp. 2067-2070, 2001.

[124] J. Lu, K.N. Plataniotis and A.N. Venetsanopoulos, “Regularized Discriminant Analysis for the Small Sample Size Problem in Face Recognition,” Pattern Recognition Letters, vol. 24, no. 16, pp. 3079-3087, 2003.

[125] W. Zheng, Li Zhao and C. Zou, “An Efficient Algorithm to Solve the Small Sample Size Problem for LDA,” *Pattern Recognition*, vol. 37, no. 5, pp. 1077-1079, 2004.

[126] A.M. Martinez and A.C. Kak, “PCA Versus LDA,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228-233, 2001.

[127] B.S. Manjunath and W.Y. Ma, “Texture Feature for Browing and Retrieval of Image Data,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837-842, 1996.

[128] C. Liu and H. Wechsler, “Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition,” *IEEE Transactions Image Processing*, vol. 11, no. 4, pp. 467-476, 2002.

[129] D. Liu, K.M. Lam, and L.S. Shen, “Optimal Sampling of Gabor Features for Face Recognition,” *Pattern Recognition Letters*, vol. 25, no. 2, pp. 267-276, 2004.

[130] T.S. Lee, “Image Representation Using 2D Gabor Wavelets,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 959-971, 1996.

[131] L. Shams and C. Malsburg, “The Role of Complex Cells in Object Recognition,” *Vision Research*, vol. 42, no. 22, pp. 2547-2554, 2002.

[132] B. Takács, “Comparing Face Images Using the Modified Hausdorff Distance,” *Pattern Recognition*, vol. 31, no. 12, pp. 1873-1880, 1998.

[133] Yale University [Online]. Available: http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html

[134] P.W. Hallinan, "A Low-Dimensional Representation of Human Faces for Arbitrary Lighting Conditions," *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*, pp. 995-999, 1994.

[135] M. Bichsel, "Illumination Invariant Object Recognition," *Proceedings of the International Conference on Image Processing*, vol. 3, pp. 620-623, 1995.

[136] A. Shashua, "On Photometric Issues in 3D Visual Recognition from a Single 2D Image," *International Journal of Computer Vision*, vol. 21, pp. 99-122, 1997.

[137] A.S. Georghiades, D.J. Kriegman, P.N. Belhumeur, "Illumination Cones for Recognition Under Variable Lighting Faces," *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*, pp. 52-58, 1998.

[138] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, "From Few to Many: Generative Models for Recognition Under Variable Pose and Illumination," *Proceedings of IEEE Conference Face and Guesture Recognition*, pp. 277-284, 2000.

[139] R. Ishiyama, S. Sakamoto, "Geodesic Illumination Basis: Compensating for Illumination Variations in Any Pose for Face Recognition," *Proceedings of the* 16th *International Conference Pattern Recognition*, vol. 4, pp. 297-301, 2002.

[140] A.U. Batur, M.H. III Hayes, "Linear Subspaces for Illumination Robust Face Recognition," *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*, vol. 2, pp. 296-301, 2001.

[141] A.Z. Kouzani, F. He, K. Sammut and A. Bouzerdoum, "Illumination Invariant Face Recognition," *Proceedings of IEEE Conference Systems, Man, and Cybernetics*, vol. 5, pp. 4240-4245, 1998.

[142] L. Zhao, Y. Yang, "Theoretical Analysis of Illumination in PCA-Based Vision Systems," *Pattern Recognition*, vol. 32, no. 4, pp. 547-564, 1999.

[143] W. Zhao and R. Chellappa, "Illumination-Insensitive Face Recognition Using Symmetric Shape-From-Shading," *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*, pp. 286-293, 2000.

[144] H.F. Chen, P.N. Belhumeur, D.W. Jacobs, "In Search of Illumination Invariants," *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*, vol.1, pp. 254-261, 2000.

[145] T. Riklin-Raviv, A. Shashua, "The Quotient Image: Class Based Recognition and Synthesis Under Varying Illumination Conditions,"

Proceedings of IEEE Conference Computer Vision and Pattern Recognition, vol. 2, pp. 23-25, 1999.

[146] A. Shashua, T. Riklin-Raviv, "The Quotient Image: Class Based Re-Rendering and Recognition with Varying Illuminations," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 129-139, 2001.

[147] D. Liu, K.M. Lam and L.S. Shen, "Illumination Invariant Face Recognition," *Pattern Recognition* accepted, 2005.

[148] C.E. Izard, "The Maximally Discriminative Facial Movement Coding System (MAX)," unpublished manuscript, Available from Instructional Resource Center, University of Delaware, Newark, Delaware, 1979.

[149] C. E. Izard, L. M. Dougherty, and E. A. Hembree, "A System for Identifying Affect Expressions by Holistic Judgments," unpublished manuscript, University of Delaware, 1983.

[150] R. Koenen, "Mpeg-4 Project Overview," International Organisation for Standardization, ISO/IEC JTC1/SC29/WG11, La Baule, 2000.

[151] N. Tsapatsoulis, K. Karpouzis, G. Stamou, "A Fuzzy System for Emotion Classification Based on the MPEG-4 Facial Definition Parameter," *European Association on Signal Processing EUSIPCO*, 2000.

[152] M. Hoch, G. Fleischmann and B. Girod, "Modeling and Animation of Facial Expressions Based on B-Splines," *Visual Computer*, vol. 11, no. 2, pp 87-95, 1994.

[153] M. Pantic, L.J.M. Rothkrantz, "Facial Action Recognition for Facial Expression Analysis from Static Face Images," *IEEE Transactions Systems, Man and Cybernetics - Part B*: *Cybernetics*, vol. 34, no. 3, pp. 1449-1461, 2004.

[154] B.A. Donohue, J.D. Bronzino, J.H. DiLiberti, D.P. Olson, L.R. Schweitzer and P. Walsh, "Application of A Neural Network in Recognizing Facial Expression," *Proceedings of the 7th IEEE Annual Northeast Bioengineering Conference*, pp. 206-207, 1991.

[155] C.S. Choi, K. Aizawa, H. Harashima, T. Takebe, "Analysis and Synthesis of Facial Image Sequences in Model-Based Image Coding," *IEEE Transactions Circuits and Systems for Video Technology*, vol. 4, no. 3, pp. 257-275, 1994.

[156] Y. Yacoob and L.S. Davis, "Recognizing Human Facial Expressions From Long Image Sequences Using Optical Flow," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, pp. 636-642, 1996.

[157] C.L. Huang, Y.M. Huang, "Facial Expression Recognition Using Model-Based Feature Extraction and Action Parameters Classification," *Journal of Visual Communication and Image Representation*, vol. 8, no. 3, pp. 278-290, 1997.

[158] N. Oliver, A. Pentland and F. Bérard, "LAFTER: A Real-Time Face and Lips Tracker with Facial Expression Recognition," *Pattern Recognition*, vol. 33, no. 8, pp. 1369-1382, 2000.

[159] Y. Zhu, L.C. De Silva, C.C. Ko, "Using Moment Invariants and HMM in Facial Expression Recognition," *Pattern Recognition Letters*, vol. 23, no. 1-3, pp. 83-91, 2002.

[160] J.N. Bassili, "Emotion Recognition: The Role of Facial Movement and the Relative Importance of Upper and Lower Areas of the Face," *Journal of Personality and Social Psychology*, vol. 37, no. 11, pp. 2049-2058, 1979.

[161] Y. Gao, M.K.H. Leung, S.C. Hui, M.W. Tananda, "Facial Expression Recognition From Line-Based Caricatures," *IEEE Transactions Systems, Man and Cybernetics - Part A*: *Systems and Humans*, vol. 33, no. 3, pp. 407-412, 2003.

[162] G. Cottrell and J. Metcalfe, "Face, Gender and Emotion Recognition Using Holons," in *Advances in Neural Information Processing Systems*, D. Touretzky, Ed., vol. 3, pp. 564–571, 1991.

[163] C. Padgett and G. Cottrell, Representing Face Images for Emotion Classification, in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, vol. 9, 1997.

[164] M.J. Lyons, J. Budynek and S. Akamatsu, "Automatic Classification of Single Facial Images," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357-1362, 1999.

[165] X. Chen and T. Huang, "Facial Expression Recognition: A Clustering-Based Approach," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1295-1302, 2003.

[166] M. Matsugu, K. Mori, Y. Mitari and Y. Kaneda, "Subject Independent Facial Expression Recognition with Robust Face Detection Using a Convolutional Neural Network," *Neural Networks*, vol. 16, no. 5-6, pp. 555-559, 2003.

[167] B. Abboud, F. Davoine and M. Dang, "Facial Expression Recognition and Synthesis Based on an Appearance Model," *Signal Processing*: *Image Communication*, vol. 19, no. 8, pp. 723-740, 2004.

[168] L. Ma and K. Khorasani, "Facial Expression Recognition Using Constructive Feedforward Neural Networks," *IEEE Transactions Systems*,

Man and Cybernetics - Part B: *Cybernetics*, vol. 34, no. 3, pp. 1588-1595, 2004.

[169] C. Liu and H. Wechsler, "A shape- and Texture-Based Enhanced Fisher Classifier for Face Recognition," *IEEE Transactions Image Processing*, vol. 10, no. 4, pp. 598-608, 2001.

[170] B.W. Hwang and S.W. Lee, "Reconstruction of Partially Damaged Face Images Based on a Morphable Face Model," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 25, no. 3, pp. 365-372, 2003.

[171] A. Lanitis, C.J. Taylor, and T.F. Cootes, "Automatic Interpretation and Coding of Face Images Using Flexible Models," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 743-765, 1997.

[172] T. Vetter and T. Poggio, "Linear Object Classes and Image Synthesis From a Single Example Image," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 733-742, 1997.

[173] K.W. Wong, K.M. Lam and W.C. Siu, "An Efficient Algorithm for Human Face Detection and Facial Feature Extraction Under Different Conditions," *Pattern Recognition*, vol. 34, no. 10, pp. 1993-2004, 2001.

[174] W.P. Choi, K.M. Lam and W.C. Siu, "An Adaptive Active Contour Model for Highly Irregular Boundaries," *Pattern Recognition*, vol. 34, pp. 323-331, 2001.

[175] Biederman, "Recognition-by-Components: A Theory of Human Image Understanding," *Psychological Review*, vol. 94, pp. 115-147, 1987.

[176] V. Bruce, E. Hanna, N. Dench, P. Healey and M. Burton, "The Importance of "mass" in Line-Drawings of Faces," *Applied Cognitive Psychology*, vol. 6, pp. 619- 628, 1992.

[177] B. Guo, K.M. Lam, W.C. Siu and S. Yang, "Human Face Recognition Based on Spatially Weighted Hausdorff Distance," *Pattern Recognition Letters*, vol. 24, no. 1-3, pp. 499-507, 2003.

[178] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, MA: Addison-Wesley, 1993.

[179] The Oliver Research Laboratory in Cambridge, UK [Online]. Available: http://www.uk.research.att.com/pub/data/att_faces.zip

[180] Y. Yitzhaky and E. Peli, "A Method for Objective Edge Detection Evaluation and Detector Parameter Selection," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 1027-1033, 2003.

[181] K.M. Lam and H. Yan, “Locating and Extracting the Eye in Human Face Images,” *Pattern Recognition*, vol. 29, no. 5, pp. 771-779, 1996.

[182] Y. Moses, “Face Recognition: Generalization to Novel Images,” PhD thesis, Weizman Inst. of Science, 1993.

[183] W. Feller, An Introduction to Probability Theory and Its Applications, vol. 2, 3rd ed., New York: Wiley, 1971.

[184] M. Abramowitz, and I.A. Stegun, editors, “Error Function and Fresnel Integrals,” in: Ch. 7 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing, New York: Dover, pp. 297-309, 1972.

[185] W. Yambor, B. Draper and R. Beveridge, Analyzing PCA-Based Face Recognition Algorithms: Eigenvector Selection and Distance Measures, in: H. Christensen and J. Phillips, editors, Empirical Evaluation Methods in Computer Vision, World Scientific Press, Singapore, 2002.

[186] S.M. Pizer, E.P. Amburn, “Adaptive Histogram Equalization and Its Variations,” *Computer Vision, Graphics, and Image Processing*, vol. 39, pp. 355-368, 1987.

[187] A.M. Vossepoel, B.C. Stoel, A.P. Meershoek, “Adaptive Histogram Equalization Using Variable Regions,” *Proceedings 9th International Conference Pattern Recognition*, vol. 1, pp. 351-353, 1988.

[188] Y. Jin, L. Fayad and A. Laine, “Contrast Enhancement by Multi-Scale Histogram Equalization,” *Proceedings of the SPIE in Wavelet Applications in Signal and Image Processing IX*, pp. 206-213, 2001.

[189] J.D. Fahnestock and R.A. Schowengerdt, “Spatially Variant Contrast Enhancement Using Local Range Modification,” Optical Engineering, vol. 22, no. 3, pp. 378-381, 1983.

[190] J. Zhu, B. Liu and S.C. Schwartz, “General Illumination Correction and Its Application to Face Normalization,” *Proceedings of IEEE Conference Acoustics, Speech, and Signal Processing*, vol. 3, pp. 133-136, 2003.

[191] A. Goshtasby, “Piecewise Cubic Mapping Functions for Image Registration,” *Pattern Recognition*, vol. 20, no. 5, pp. 525-533, 1987.

[192] W. Zhao, R. Chellappa and P.J. Phillips, “Subspace Linear Discriminant Analysis for Face Recognition,” Technical Report CAR-TR-914, CS-TR-4009, University of Maryland at College Park, USA, 1999.

[193] Yale University [Online]. Available: http://cvc.yale.edu/projects/yalefaces/yalefaces.html

[194] A.M. Martinez and R. Benavente, The AR face database, CVC Technical Report #24, June 1998.

[195] H. Moon and J. Phillips, Analysis of PCA-Based Face Recognition Algorithms, In: K. Boyer and J. Phillips, editors, Empirical Evaluation Techniques in Computer Vision. IEEE Computer Society Press, 1998.

[196] X. Xie, K.M. Lam, "An Efficient Method for Face Recognition Under Varying Illumination," *Proceedings of the International Symposium Circuits and Systems*, 2005.

[197] G.C. Feng and P.C. Yuen, "Recognition of head-and-shoulder face image using virtual frontal-view image," *IEEE Transactions Systems, Man and Cybernetics, Part A*, vol. 30, no. 6, pp. 871-882, 2000.

[198] J. Ahlberg, "CANDIDE-3 - Un Updated Parameterised Face," Report No. LiTH-ISY-R-2326, 2001.

[199] T. Sim, S. Baker and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database," *Proceedings of the IEEE Conference Automatic Face and Gesture Recognition*, 2002.

[200] K.C. Lee, J. Ho and D. Kriegman, "Nine Points of Light: Acquiring Subspaces for Face Recognition Under Variable Lighting," *Proceedings of International Conference Computer Vision and Pattern Recognition*, vol. 1, pp. 519-526, 2001.

[201] J. Hurri, H. Gävert, J. Särelä, A. Hyvärinen, Laboratory of Information and Computer Science in the Helsinki University of Technology [Online]. Available: http://www.cis.hut.fi/projects/ica/fastica/code/FastICA_2.1.tar.gz, 2004.

[202] K.H. Lin, B. Guo, K.M. Lam and W.C. Siu, "Human Face Recognition Using a Spatially Weighted Modified Hausdorff Distance," *Proceedings of 2001 International Symposium Intelligent Multimedia, Video and Speech Processing*, pp. 447-480, 2001.

[203] M. Flicker, H. Sawhney, W. Niblack, J. Ashley, H. Qian, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele and P. Yanker, "Query by Image and Video Content: The QBIC System," *Computer*, vol. 28, no. 9, pp. 23-32, 1995.

[204] J.R. Bach, "The Virage Image Search Engine: An Open Framework for Image Management," *Proceedings of Conference Storage and Retrieval for Image and Video Databases IV*, vol. 2670, pp.76-87, 1996.

[205] A.K. Jain and A. Vailaya, "Image Retrieval Using Color and Shape," *Pattern Recognition*, vol. 29, no. 8, pp. 1233-1244, 1996.

[206] M. Kliot and E. Rivlin, "Invariant-Based Shape Retrieval in Pictorial Databases," *Computer Vision and Image Understanding*, vol. 71, no. 2, pp.182-197, 1998.

[207] J. Vleugels and R.C. Veltkamp, "Efficient Image Retrieval Through Vantage Object," *Pattern Recognition*, vol. 35, no. 1, pp. 69-80, 2002.

[208] B. Weyrauch, J. Huang, B. Heisele, and V. Blanz. "Component-Based Face Recognition with 3D Morphable Models," *First IEEE Workshop on Face Processing in Video*, Washington, D.C., 2004.

[209] O. Jesorsky, K. Kirchberg and R. Frischholz, Robust Face Detection Using the Hausdorff Distance, In J. Bigun and F. Smeraldi, editors, Audio and Video based Person Authentication - AVBPA 2001, pp. 90-95. Springer, 2001.

[210] D.B. Graham and G.N.M. Allinson, "Characterizing Virtual Eigensignatures for General Purpose Face Recognition," In *Face recognition*: *From theory to applications*, NATO ASI Series F, Computer and Systems Sciences 163, 446-456, 1998.

[211] Bern University face database, [Online]. Available: ftp://iamftp.unibe.ch/pub/Images/FaceImages/

[212] D. Voth, "Face Recognition Technology," *Intelligent Systems, IEEE* [see also *IEEE Expert*], vol. 18, no. 3, pp. 4-7, May-June 2003.

[213] C. Zhang and S.C. Fernand, "3-D Face Structure Extraction and Recognition From Images Using 3-D Morphing and Distance Mapping," *IEEE Transactions Image Processing*, vol. 11, no. 11, pp. 1249-1259, 2002.