Face recognition using soft computing methods

Chen, Weilong

2006

Chen, W. (2006). Face recognition using soft computing methods. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/39140>

<https://doi.org/10.32657/10356/39140>

a 455865

Face Recognition Using Soft Computing Methods

Chen Weilong

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in fulfilment of the requirement for the degree of
Doctor of Philosophy

2006

Statement of Originality

I hereby certify that the content of this thesis is the result of work done by me and has not been submitted for a higher degree to any other University or Institution.

Oct. 19, 2006

Date

Chen Weilong

Chen Weilong

Acknowledgments

I would like to take this opportunity to express my sincere appreciation to many people who have helped me in one way or another during my doctoral study at Nanyang Technological University.

First and foremost, I wish to express my deepest gratitude to my supervisor, Associate Professor Er Meng Joo, for his guidance, support and encouragement which make the completion of the thesis possible. I am also very grateful to Dr. Wu Shiqian for his valuable advice on my research work.

I would like to thank all my colleagues and friends in the School of Electrical and Electronic Engineering for their constant assistance.

I am greatly indebted to my parents for their support and understanding. My special gratitude should go to my beloved wife, Wang Hui. Without her love, support and encouragement, my thesis would not have been successfully finished.

Summary

Human face recognition is still a challenging task for computer-based machines mainly due to the similarity of human faces and unpredictable variations such as illumination, pose, expression, occlusion etc. A lot of algorithms have been developed to deal with these challenging issues in face recognition. One of the difficult problems is the geometric variation of face images. In some face recognition applications, only face images with uncontrolled format are available. Face images obtained from face detection stage may be directly used. Therefore, alignment or normalization is required to achieve better recognition performance. Eye detection technique is normally used to normalize face images. Invariant representations are used to improve the generalization of facial features. Feature-based and model-based automatic face alignment can also be used to localize human faces more precisely.

Illumination variation is another tough problem because human face is a 3D model and practical lighting conditions are complicated and unpredictable. Illumination normalization based on preprocessing, illumination invariant feature extraction and face modelling have been employed to reduce illumination variations or render faces under varying lighting conditions. The geometric on-the-plane rotation of a face image can be normalized by aforementioned methods. However, a face image with rotation in depth is not easily adjusted to a certain view. It is difficult to recognize faces in all poses with only training face images in limited views. One solution is to use 2D or 3D generative models to generate novel views of the face. Another method is based on the view-subspace learning. For facial expression variations,

a few approaches on variation discounting and face expression synthesizing have also been proposed.

In this thesis, some soft computing methods for automatic face recognition under variable circumstances are explored and investigated. First, we prove that statistical feature extraction approaches such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) can be directly implemented in the Discrete Cosine Transform (DCT) domain. Therefore, LDA can be implemented in the truncated DCT domain to obtain discriminating features as well as to achieve computational efficiency. Based on the theoretical proof, experimental results also show that PCA, LDA and D-LDA in the truncated DCT domain achieve similar performance with the one in the original domain. In the classification stage, RBF neural networks are employed because they possess a few advantages over the conventional multilayered networks. Combined with the proposed clustering algorithms and parameter estimation methods, good recognition performance as well as high training and recognition speed are achieved.

A novel illumination normalization approach is proposed in this thesis. For face images, illumination variations usually change slowly compared with the reflectance of facial features. Based on the reflectance theory, illumination variations under different lighting conditions can be significantly reduced by discarding low-frequency DCT coefficients in the logarithm domain. One of the advantages of this approach is that it is very fast and can be easily implemented in a real-time face recognition system because the DCT can normally implemented by a fast algorithm. Experimental results show that the proposed approach improves the performance significantly for face images with large illumination variations.

In order to improve generalization of the facial feature extraction, a third-order invariant representation based on gradient images is proposed to achieve practical translation, scale and rotation invariance. Different from some other applications for triple-correlation-based neural networks, gray-level information of gradient face images is directly used to keep the most important facial features. The main ad-

vantage of the third-order invariant representation is that higher-frequency features can be efficiently extracted without alignment procedure. Furthermore, combined with the proposed illumination normalization approach, robustness against illumination variations can also be improved even for mis-aligned face images.

Table of Contents

Acknowledgments	i
Summary	ii
List of Figures	x
List of Tables	xv
1 Introduction	1
1.1 Motivation	1
1.2 Objective	5
1.3 Major Contributions of the Thesis	6
1.4 Organization of the Thesis	7
2 A Literature Review on Face Recognition	8
2.1 Appearance-based Face Recognition	8
2.1.1 Eigenface	9
2.1.2 Fisherface	11
2.1.3 Independent Component Analysis	13
2.1.4 Nonlinear Subspace Analysis	15

TABLE OF CONTENTS	vi
2.1.5 Edge Map	17
2.1.6 Hidden Markov Models	19
2.2 Feature-based Face Recognition	21
2.2.1 Dynamic Links	22
2.2.2 Elastic Bunch Graph Matching	24
2.2.3 Active Appearance Model	25
2.3 Neural Networks	29
2.4 Face Databases	31
3 PCA and LDA in Discrete Cosine Transform Domain	33
3.1 Introduction	33
3.2 PCA in DCT Domain	35
3.2.1 PCA on Orthonormally Transformed Data	35
3.2.2 Discrete Cosine Transform	37
3.2.3 Block DCT	39
3.3 LDA in DCT Domain	40
3.4 D-LDA in DCT Domain	42
3.5 Experimental Results	44
3.6 Conclusion	47
4 High-Speed Face Recognition Based on Discrete Cosine Transform and RBF Neural Networks	52
4.1 Introduction	53
4.2 Feature Extraction	55
4.2.1 Discrete Cosine Transform (DCT)	55

LIST OF FIGURES

xi

3.4	Performance of PCA in JPEG DCT domain (Dup 1).	48
3.5	Performance of PCA in JPEG DCT domain (Dup 2).	48
3.6	Performance of PCA in JPEG DCT domain (Fafb).	49
3.7	Performance of PCA in JPEG DCT domain (Fafc).	49
3.8	Performance of D-LDA in JPEG DCT domain (Dup 1).	50
3.9	Performance of D-LDA in JPEG DCT domain (Dup 2).	50
3.10	Performance of D-LDA in JPEG DCT domain (Fafb).	51
3.11	Performance of D-LDA in JPEG DCT domain (Fafc).	51
4.1	Block diagram of high-speed face recognition based on DCT and RBF neural networks	55
4.2	Scheme of scanning 2D DCT coefficients to a one-dimensional vector	56
4.3	Clustering with category information	58
4.4	Illustration of one class split into three subclasses	60
4.5	Example of training samples in one class splitted into two subclasses by the clustering algorithm ($\alpha = 1$)	61
4.6	Architecture of RBF neural networks	62
4.7	The ORL face database	68
4.8	RMSE curve	71
4.9	Performances with different number of training samples (Results are based on ten runs)	76
4.10	Examples of the centered Yale images	76
4.11	Examples of the closely cropped Yale images	77
5.1	DCT basis images.	86

5.2 (a) Original image; (b) Reconstructed image by applying DCT on the original image and discarding the (0, 1)th DCT coefficient; (c) Reconstructed image by applying DCT on the logarithm image and discarding the (0, 1)th DCT coefficient ($\mu = 100$); (d) Reconstructed logarithm image by applying DCT on the logarithm image and discarding the (0, 1) DCT coefficient (i.e., (c) without the inverse logarithm transform). 86

5.3 Standard deviations of the logarithm DCT coefficients. 88

5.4 Manner of discarding DCT coefficients. 88

5.5 Sample images of an individual divided into 5 subsets. 90

5.6 Sample images of an individual in CMU PIE database: (a) Training; (b) Testing. 91

5.7 Normalized logarithm images with different D_{dis} : (a)Original image; (b) $D_{dis} = 3$; (c) $D_{dis} = 6$; (d) $D_{dis} = 15$; (e) $D_{dis} = 20$; (f) $D_{dis} = 35$; (g) $D_{dis} = 50$ 92

5.8 Performance on the Yale B database with different D_{dis} (Correlation). 93

5.9 Performance on the Yale B database with different D_{dis} (Eigenfaces). 93

5.10 Performance on the CMU PIE database with different D_{dis} 94

5.11 Performance comparison between logarithm and non-logarithm images (Yale B Subset 4). 95

5.12 Performance comparison between logarithm and non-logarithm images (Yale B Subset 5). 96

5.13 Performance comparison between logarithm and non-logarithm images (CMU PIE). 97

5.14 Performance based on the Eigenface method by discarding different numbers of principal components (50 principal components are used, i.e., the dimension of feature vectors is 50). 98

LIST OF FIGURES	xiii
5.15 Performance on the Yale B database using high-pass filters with different transfer functions (Subset 4).	98
5.16 Performance on the Yale B database using high-pass filters with different transfer functions (Subset 5).	99
5.17 Performance on the CMU PIE database using high-pass filters with different transfer functions.	100
5.18 (a) 2nd-order Butterworth filter; (b) 4th-order Butterworth filter; (c) Ideal filter. (cutoff = 0.06)	100
5.19 Average variation estimation efficiency (VEE) of the Yale B database (Coefficients are sorted according to the DCT discarding and DFT filtering manner).	101
5.20 Average variation estimation efficiency (VEE) of the CMU PIE database (Coefficients are sorted according to the DCT discarding and DFT filtering manner).	102
5.21 Examples of mis-aligned face images from Yale B Subset 1.	102
5.22 Performance on the mis-aligned Yale B database with different D_{dis} (Correlation).	103
5.23 Performance on the mis-aligned Yale B database with different D_{dis} (Eigenfaces).	104
5.24 Performance on the mis-aligned CMU PIE database with different D_{dis}	104
6.1 Example of face images with ideal geometric variations	109
6.2 Example of face images with practical geometric variations	109
6.3 Feature extraction based on Sobel operator: (a) Original image; (b) Sobel gradient image; (c) Thresholded binary image.	110

6.4	Feature extraction on logarithm image: (a) Logarithm image; (b) Sobel gradient image; (c) Thresholded binary image.	111
6.5	Extracted feature images after morphological operations: (a) Binary image after morphological operations; (b) Extracted Sobel gradient image; (c) Extracted logarithm gray-level image.	111
6.6	Third-order neural network with 4 inputs and 1 output.	112
6.7	Built-in invariance of face images with translation, scale and rotation variations	113
6.8	Invariant image space with respect to angles a and b	115
6.9	Example of partial connectivity.	116
6.10	Example of face images used for training and testing: (a) Training; (b) Testing (Subset 1).	117
6.11	Higher-order invariant representations of gradient images with translation, scale and rotation variations ($D_b = 20$, $\alpha = 0.5^\circ$).	119
6.12	Performance with different partial connectivity distance bound D_b ($\alpha = 0.5$).	120
6.13	Higher-order invariant representations with different quantization degrees ($D_b \geq 20$).	121
6.14	Performance with different quantization degree α ($D_b = 20$).	122
6.15	Example of mis-aligned face images with illumination variations from Yale B Subset 3.	122
6.16	Performance with different D_{dis} based on higher-order representations.	123
6.17	Example of a face image with shadowing variations and the extracted gradient feature image.	123

List of Tables

2.1	Face Databases	32
3.1	Four probe subsets and their evaluation task	45
4.1	Recognition performance versus number of DCT coefficients ($\alpha = 1$, $\gamma = 0.1$)	71
4.2	Recognition performance versus clustering factor α ($\gamma = 0.1$)	72
4.3	Performance on 10 simulations ($\alpha = 1.5$, $\gamma = 0.1$)	73
4.4	Recognition performance comparison of different approaches	74
4.5	Computational complexity	75
4.6	Comparison on the Yale database	77
4.7	Performance on the Yale database ($\gamma = 0.1$)	78
5.1	Subsets divided according to light source direction	89
5.2	Recognition performance comparisons of different methods	94
5.3	Recognition performance comparison between DCT and DFT (Correlation)	100
5.4	Recognition performance comparison between well-aligned and mis-aligned images (Correlation, $18 \leq D_{dis} \leq 25$)	101

6.1	Feature extraction comparisons between original image and logarithm image ($D_b = 20, \alpha = 0.5^\circ$).	118
6.2	Results of the eigenface approach on face images with geometric variations.	118
6.3	Number of values in representations with different α	122
6.4	Performance on the ORL face database ($\alpha = 0.5, D_b = 20$).	124

TABLE OF CONTENTS		vii
4.2.2	Clustering	57
4.2.3	Linear Discriminant Analysis (LDA)	61
4.3	Classification Using RBF Neural Networks	62
4.3.1	RBF Neural Networks	62
4.3.2	Structure Determination and Parameter Estimation of RBF Neural Networks	63
4.4	Experimental Results and Discussion	67
4.4.1	Testing on the ORL Database	67
4.4.2	Testing on the Yale Database	75
4.5	Conclusion	77
5 Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithm Do- main		79
5.1	Introduction	80
5.2	Illumination Normalization in the Logarithm DCT Domain	82
5.2.1	Logarithm Transform	82
5.2.2	Illumination Compensation	83
5.2.3	Logarithm Domain versus Original Domain	86
5.2.4	Logarithm Image for Recognition	87
5.2.5	Discarding DCT Coefficients	87
5.3	Experimental Results and Discussion	89
5.3.1	Face Database	89
5.3.2	Experimental Results	90

TABLE OF CONTENTS

viii

5.3.3	Performance Comparison between Logarithm and Non-logarithm Images	95
5.3.4	Discarding DCT Coefficients versus Discarding PCA Components	95
5.3.5	DCT versus DFT	96
5.3.6	Performance with Mis-Aligned Face Images	100
5.4	Conclusions	102
6	Higher-Order Invariant Representation for Face Recognition	105
6.1	Introduction	105
6.2	Preprocessing and Feature Extraction	108
6.2.1	Gradient Image	109
6.2.2	Illumination and Noise Reduction	110
6.3	Third-Order Neural Networks	112
6.4	Higher-Order Invariant Representation	114
6.4.1	Partial Connectivity Strategies	115
6.5	Experimental Results	116
6.5.1	Original Image versus Logarithm Image	117
6.5.2	Effect of Partial Connectivity Strategy	118
6.5.3	Performance with Different Quantization Degrees of Included Angles	120
6.5.4	Performance with DCT Illumination Normalization in Logarithm Domain	121
6.5.5	Performance on the ORL Face Database	123
6.6	Conclusions	124

TABLE OF CONTENTS

ix

7	Conclusions and Recommendations for Future Work	126
7.1	Conclusions	126
7.2	Recommendations for Future Work	128
	Author’s Publications	131
	Bibliography	133

List of Figures

2.1 *Eigenfaces* of ORL face database: (a) Average image; (b) *Eigenfaces* corresponding to the 10 largest eigenvalues, ordered left to right, top to bottom. 10

2.2 Comparison between PCA and LDA projection [25]. 12

2.3 Comparison between PCA and ICA projection [25]. 14

2.4 Basic idea of kernel PCA. 15

2.5 Top-bottom 5 state HMM [42]. 21

2.6 The shape of a wavelet (The real part and the imaginary part) [50]. 22

2.7 Example of matching [50]. 23

2.8 Labelled graph with different poses [53]. 26

2.9 Face bunch graph (FBG) [53]. 26

2.10 A labelled training image gives a shape free patch and a set of points [56]. 27

2.11 Multiresolution search from displaced position using face model [56]. 28

3.1 Examples of normalized FERET face images from four subsets: (a) Dup 1; (b) Dup 2; (c) Fafb; (d) Fafc 46

3.2 Feature vector formation in block DCT domain. 47

3.3 A typical quantization matrix. 47

Chapter 1

Introduction

1.1 Motivation

In recent years, there has been a growing interest in machine recognition of faces due to its potential commercial and law enforcement applications. These applications range from static matching of controlled format images such as credit cards, passports, driver's licenses and mug shots to dynamic matching such as crowd surveillance, bank/store security and access control systems etc., i.e. image-based and video-based face recognition [1]. Different applications involve different kinds of issues and requirements. Compared with other biometric recognition methods, such as fingerprint, voice, retina, iris and palm geometry etc., face recognition technique possesses more convenience because it is non-intrusive and requires little cooperation of participants.

More specifically, there are two types of applications related to face recognition techniques: (1) face verification (or authentication); (2) face identification (or recognition). Verification is a one-to-one matching process in which a query face image is matched against a known face image in a face database to determine whether the user is who he claims to be. The computational cost is independent of the number of recorded users in the database. The evaluation of verification

performance is mainly based on False Rejection Rate (FRR), False Acceptance Rate (FAR) and Equal Error Rate (EER). The Receiver Operating Characteristics (ROC) curve is also employed to evaluate the trade-off between the FRR and FAR [2, 3]. Identification is a one-to-many match that compares a query face image against all the face images in a face database in order to determine the identity of the query face image. The computational cost for the identification increases in proportion to the number of face images in the database. Identification rates (or error rates) can be simply used to evaluate a face identification system. Some face recognition evaluation systems (e.g. FERET) employ Cumulative Match Score to evaluate the identification performance [4, 5]. In this thesis, we mainly focus on face identification.

Generally speaking, a complete face recognition system should include three steps:

- (1) **Face detection.** This step is also called as localization or segmentation of faces in some literature. In some applications, face images for recognition are obtained under controlled conditions, e.g. the passport photographs taken with plain background and approximately in the same scale. Therefore, the location of a face can be easily determined. However, in most cases, the location of a face in the image is unknown. The goal of face detection is to find out whether or not there are any faces in a given image and determine the location and extent of each face in the image [6, 7]. After finding the face, we can segment the face to count out useless information of the image such as variable backgrounds. Face detection encounters several challenges such as variations in pose, expression, occlusion, rotation, scale and illumination.
- (2) **Feature extraction.** Feature extraction is an essential step in a face recognition system. The face recognition procedure must be based on various features of a human face. In order to get better recognition performance, it is desirable to have a simple and reliable representation of the input face image (eliminating redundant information and extracting the most salient and invariant features). The image features can be divided into four groups,

namely visual features, statistical pixel features, transform coefficient features, and algebraic features [1]. Additionally, feature extraction methods are broadly classified into two categories, namely feature-based and appearance-based (including template matching). Feature-based approaches are based on shapes and geometrical relationships of individual facial features including eyes, mouth, nose and chin. On the other hand, appearance-based approaches handle the input face images globally and automatically extract important facial features based on the high dimensional intensity values of face images. It should be noted that since some feature extraction methods are sensitive to variations such as illumination, rotation and scale, face images should be preprocessed before implementing the feature extraction step. This procedure is called normalization. However, it is usually computational expensive and based on the technique of feature detection. Therefore, it is more practical to develop an efficient feature extraction method which is invariant to changes of illumination, scale, translation, pose etc.

- (3) **Recognition (Classification).** In this step, the aim is to identify the query face image based on the images in the database. In the feature extraction stage, face images are normally converted to low dimensional feature vectors which well represent salient features of the face images. The feature vectors in the spanned feature space are then used for classification instead of the entire face images. Several efficient approaches have been proposed for pattern classification, e.g. Nearest Neighbor (NN), k -Nearest Neighbor (KNN), Nearest Center (NC), Neural Networks (NN)(e.g. Multi-layer perceptron (MLP) and Radial Basis Function (RBF) Neural Networks, Support Vector Machine (SVM)) [8, 9], Nearest Feature Line (NFL) [10] etc.

A great number of face recognition approaches have been proposed and modern face recognition has reached a high recognition rate under constrained conditions. However, face recognition systems with 100% reliable recognition rate are desirable in most practical applications. Moreover, in some applications, face images

are not always with well-controlled conditions. Instead, they are with variations in illumination, pose and facial expression, etc. Therefore, highly accurate face recognition under variable circumstances is still a challenging task for computer-based machines. Today's face recognition research mostly focuses on the following challenging issues:

- **Illumination variations.** Illumination variation conditions can be divided into three cases: (1) Uniform illumination with linear scale of lighting intensity; (2) Uniform illumination with nonlinear scale of lighting intensity; (3) Non-uniform illumination. Uniform illumination variations can be alleviated by using image processing techniques. The most difficult one to deal with is the non-uniform illumination in which the light source may come from different directions and can form various regions of shadowing and reflection on human faces. Illumination invariant approaches can be briefly categorized into three groups: (1) Preprocess face images to remove various unwanted illumination effects; (2) Employ efficient feature extraction methods which can ignore illumination variations; (3) Construct face models which can be used to render face images under new illumination conditions.
- **Pose variations.** In some applications, a face recognition system must be able to recognize faces with different poses. The orientation of a face image, or the on-the-plane rotation of a face image, can be rectified by a normalization process together with scale and position. However, face images with rotation in depth are totally different and not easily adjusted to certain views. One strategy for pose invariant recognition is to create a dataset with different pose views for each person. Since in some cases it is impossible to get enough training samples in different views, some approaches have been proposed to generate new views from limited training samples. Another solution for pose variations is to employ 3D images for recognition which are robust against changes in pose.
- **Expression variations.** Human face is a non-rigid object. Facial expression

variations also bring about difficulties for precisely recognizing faces. Like the solution to pose variations, increasing training samples with different expressions may improve the performance. More efficiently, for the case with limited training samples, some researchers have proposed a few approaches based on face models or non-rigid face matching to deal with expression variations.

- Occlusion. A face may be partially blocked by something, mainly glasses, scarves, etc. In this situation, holistic feature extraction methods may fail. Therefore, local features should be employed instead to recognize faces in order to discount the occluded parts. This situation makes the recognition task more difficult since some useful facial features may not be used.
- Real-time performance. Assessment of face recognition systems is usually based on three aspects: 1) Accuracy; 2) Robustness; 3) Speed. Although some efficient algorithms for face recognition are proposed, their computational burden is too heavy to recognize faces in real-time. High-speed recognition is desirable in some applications, e.g. in video surveillance systems. Moreover, in some cases, we need to recognize a face from a huge face database. Therefore, real-time face recognition is also a crucial problem which determines if it can be applied in some practical systems.

1.2 Objective

Our research is mainly focused on the last two steps mentioned in Section 1.1: feature extraction and classification. The main objective of this thesis is to explore some feature extraction approaches with good generalization which are robust against some of the aforementioned variations as well as to achieve computational efficiency. Since face images usually lie in complex nonlinear feature spaces, efficient classification approaches should also be developed in order to improve the performance of a face recognition system.

1.3 Major Contributions of the Thesis

- (1) In this thesis, it is shown that Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) can be directly implemented in Discrete Cosine Transform (DCT) domain and the results are exactly the same as the one obtained from the spatial domain. Therefore, for images compressed using the DCT, e.g. in JPEG or MPEG standard, PCA and LDA can be directly implemented in the DCT domain such that the inverse DCT transform can be skipped and the dimensionality of the original data can be initially reduced to reduce computational cost.
- (2) An efficient method for high-speed face recognition based on Discrete Cosine Transform (DCT), Linear Discriminant Analysis (LDA) and RBF neural networks is developed. LDA is implemented in the truncated DCT domain to achieve computational efficiency and extract the most discriminating and invariant facial features. Before implementing LDA, a clustering algorithm is proposed to improve the performance of feature extraction based on LDA. Therefore, parameter estimation for the RBF neural networks can be fulfilled easily which facilitates fast training in the RBF neural networks.
- (3) A novel illumination normalization approach for face recognition under varying lighting conditions is proposed. In the proposed approach, Discrete Cosine Transform (DCT) is employed to compensate for illumination variations in the logarithm domain. Experimental results on databases with large illumination variations show that the proposed approach improves the performance significantly for face images with large illumination variations. Moreover, the advantage of our approach is that it does not require any modelling steps and can be easily implemented in a real-time face recognition system.
- (4) A higher-order invariant representation for a face image is proposed to achieve translation, scale and rotation invariance. Thus, the freedom of image space is largely reduced to obtain more generalization ability and the face align-

ment procedure, which is a non-trivial and unreliable task, is not needed any more. Combined with the proposed illumination normalization approach, robustness against illumination variations can also be improved.

1.4 Organization of the Thesis

The rest of this thesis is organized as follows: In Chapter 2, we make a brief literature review about recent developments in the area of face recognition. Chapter 3 presents the proof of PCA and LDA in the DCT domain. A high-speed face recognition system based on the DCT and RBF neural networks is described in Chapter 4. In Chapter 5, a novel illumination normalization approach based on the DCT in logarithm domain is proposed. In Chapter 6, a higher-order invariant representation for invariant face recognition is investigated and discussed. Finally, conclusions are drawn and recommendations for future work are described in Chapter 7.

Chapter 2

A Literature Review on Face Recognition

Face recognition is a multi-disciplinary research area spanning image processing, pattern recognition, neural networks and computer vision [1]. A variety of approaches have been proposed for face recognition over the past 30 years. In this chapter, a review of well-known face recognition approaches is presented.

2.1 Appearance-based Face Recognition

For feature extraction, there are two broad categories: appearance-based (holistic) and feature-based (analytic). Hybrid face recognition techniques combining these two kinds of approaches are also presented by some researchers. In some literature, face recognition approaches may also be classified into appearance-based and model-based [11]. Model-based approaches include approaches based on 3D morphable models. In this thesis, face recognition based on 3D models will not be discussed. Appearance-based approaches are based on global facial features, considering an image as a high dimensional vector. Statistical space analysis methods are usually employed to determine the distribution of high dimensional vectors.

To obtain an efficient and simple representation as well as to remove redundant information, appearance-based approaches transform high dimensional images into a low dimensional subspace (feature space). For the holistic appearance-based approaches, there is no need to detect facial features such that they are more computationally efficient. However, appearance-based approaches are more sensitive to variations such as pose, expression, rotation, translation and scale. Accordingly, face images should be well normalized and aligned. Otherwise, much more training samples with possible variations are required for precise recognition. Nevertheless, it is impossible for some variations to be well normalized such as pose and expression. As a consequence, subspace learning is essential for appearance-based face recognition.

2.1.1 Eigenface

Eigenface is a successful appearance-based approach for face recognition. Kirby and Sirovich [12] originally applied the Karhunen-Loeve transform (KLT) (the equivalent Principal Component Analysis (PCA)) to face recognition. In [13], the authors proposed a face recognition system based on the PCA in which face images are projected onto principal components of the original set of training face images. The principal components, or the eigenvectors of the covariance matrix of the set of face images can be displayed as ghostly faces which are called *eigenfaces*. Fig. 2.1 shows the average face image and the *eigenfaces* corresponding to the 10 largest eigenvalues from the ORL face database. Face images can be approximately reconstructed by a weighted sum of significant *eigenfaces*. Therefore, low dimensional weights can be used to represent the original face images for recognition. Given a set of N n -dimensional training images $\{x_1, x_2, \dots, x_N\}$, the covariance matrix is defined as

$$C_x = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T = XX^T \quad (2.1)$$

where $\bar{x}$ is the mean of N training samples

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.2)$$

The PCA projection matrix A is composed of m ($m < n$) eigenvectors corresponding to the first m largest eigenvalues of the total scatter matrix. We can simply obtain the m -dimensional subspace (i.e. feature space) representation of the i th image by

$$y_i = A^T(x_i - \bar{x}) \quad (2.3)$$

(a)

(b)

Figure 2.1: *Eigenfaces* of ORL face database: (a) Average image; (b) *Eigenfaces* corresponding to the 10 largest eigenvalues, ordered left to right, top to bottom.

In practical applications, the dimensionality of the image vectors n is very high such that the size of the total scatter matrix is too large to be computed and stored. The

solution is to calculate the largest eigenvalue and the corresponding eigenvector from the matrix $X^T X \in \mathbb{R}^{N \times N}$ instead of $XX^T \in \mathbb{R}^{n \times n}$ because normally the number of training samples N is far smaller than the image dimensionality n . The eigenface method is considered the optimal one for Gaussian distributed data since truncated reconstructions have a minimum mean square error (MSE). However, as a holistic approach, it is sensitive to variations in face orientation, position, and illumination. This limitation can be alleviated by additional preprocessing and normalization of the face images. The eigenface approach has become a key idea in the area of face recognition and a number of modified PCA approaches have been developed [14, 15, 16]. In [15], the mixture-of-eigenfaces method was proposed, which uses more than one set of features and outperforms the eigenface method for both the light-variant face image set and the pose-variant face image set. In [16], another extension of PCA, multiresolution principal component analysis (M-PCA) approach was proposed. The M-PCA decomposition obtains multiresolution features of face images. Local Feature Analysis (LFA) is also based on PCA [17].

2.1.2 Fisherface

In [18], the Fisherface approach is proposed, which is based on Fisher's Linear Discriminant (FLD) (also called Linear Discriminant Analysis (LDA)) technique. The key idea of this approach is to linearly project the high dimensional image to a low dimensional subspace which discounts those regions of the face with large deviations in illumination and facial expression. The eigenface method linearly projects the image space into a low dimensional feature space which maximizes the total scatter. However, the Fisherface approach finds an optimal subspace in which the ratio of the between-class scatter and the within-class scatter is maximized. Fig. 2.2 shows the difference between PCA and LDA. It is evident that LDA outperforms PCA in terms of class separation. The between-class scatter matrix

Figure 2.2: Comparison between PCA and LDA projection [25].

S_b and the within-class scatter matrix S_w are respectively defined as

$$S_b = \sum_{i=1}^c n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad (2.4)$$

$$S_w = \sum_{i=1}^c \sum_{x_j \in c_i} (x_j - \bar{x}_i)(x_j - \bar{x}_i)^T \quad (2.5)$$

where c is the number of classes, and n_i is the number of samples in class c_i , and $\bar{x}_i = (1/n_i) \sum_{k=1}^{n_i} x_k$ is the mean of class c_i , and $\bar{x} = (1/c) \sum_{k=1}^c \bar{x}_k$ is the mean of all samples. The optimal subspace projection is determined as follows:

$$E_{opt} = \arg \max_E \frac{|E^T S_b E|}{|E^T S_w E|} = [e_1, e_2, \dots, e_m] \quad (2.6)$$

where $[e_1, e_2, \dots, e_m]$ is the set of generalized eigenvectors corresponding to the m ($m \leq c - 1$) largest generalized eigenvalues λ_i , $i = 1, 2, \dots, m$, i.e.

$$S_b e_i = \lambda_i S_w e_i, \quad i = 1, 2, \dots, m \quad (2.7)$$

So, the i th fisher subspace representation is given by

$$y_i = E_{opt}^T x_i \quad (2.8)$$

It has been shown that the Fisherface approach is insensitive to large variations in lighting direction and facial expression. In [19], it was indicated that the Fisherface approach works efficiently only when the number of training samples is large and representative for each class. The authors have proven that when the training dataset is small, PCA can outperform LDA and PCA is less sensitive to different training datasets. Some other improved LDA-based feature extraction methods are presented in [20, 21, 22]. For LDA in face recognition, there will be singular problems in scatter matrices due to high dimensional data and small number of training samples. As a consequence, PCA + LDA (Fisherface) scheme is commonly used. Some variants (e.g. the direct LDA (D-LDA)) of LDA have been proposed to solve the small sample size problem [20, 21]. The key idea of the D-LDA is to discard the null space of S_b rather than the null space of S_w to avoid the loss of discriminating information. In [23], another improved LDA method called fractional step LDA (F-LDA) was presented. For the F-LDA method, dimensionality reduction is implemented in fractional steps in order to select weighting functions to achieve classification with high accuracy. In order to adopt F-LDA in high dimensional data applications such as face recognition, the DF-LDA method was proposed in [24] by combining the D-LDA and F-LDA methods.

2.1.3 Independent Component Analysis

More recently, Independent Component Analysis (ICA) has been applied in face recognition [25, 26, 27, 28]. In contrast with PCA which is used to find uncorrelated orthogonal basis images based on second-order statistics only, ICA decorrelates both second-order and higher-order statistics to minimize the dependency in input data. Therefore, ICA can be regarded as a generalization of PCA. Fig. 2.3 shows the differences between PCA and ICA projection for 3-D data distribution. In face

recognition, some important facial features may lie in higher-order relationships among image pixels because there are a lot of variations in illumination, expression and pose. Moreover, ICA is more related to local descriptions of face images which are very useful for face recognition since human face is a non-rigid object. The local features are less sensitive to variations and occlusion. The simplest form of ICA is described as follows. Let $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ be an n -dimensional observed random variable and let $\mathbf{s} = [s_1, s_2, \dots, s_n]^T$ be the n independent components (sources) with zero mean. Then, the mixing model of ICA is represented by

$$\mathbf{x} = A\mathbf{s} \quad (2.9)$$

where A is an invertible unknown mixing matrix. The goal of ICA is to find the mixing matrix A . Although there is no close-form solution to finding a mixing matrix A , a number of algorithms have been proposed to perform ICA such as InfoMax, JADE and FastICA [28, 29, 30]. These algorithms try to iteratively approximate A such that independence can be maximized.

Figure 2.3: Comparison between PCA and ICA projection [25].

2.1.4 Nonlinear Subspace Analysis

In previous sections, linear subspace mapping approaches are discussed. For face images, there are a lot of non-linear variations. Therefore, non-linear subspace analysis is necessary to explore the intrinsic non-linear relationship. Some manifold learning approaches have been proposed to deal with this problem, such as ISOMAP [31, 32] and Locally Linear Embedding (LLE) [33]. The ISOMAP approach is build on the classical multidimensional scaling (MDS) but seeks to preserve the intrinsic geometric of data, as captured in the geodesic manifold distance between all pairs of data points. The LLE approach eliminates the need to estimate pairwise distances between widely separated data points. Instead, it recovers global nonlinear structure from locally linear fits. Kernel-based non-linear mapping approaches have also attracted a lot of attentions in pattern recognition research [15, 34, 35, 36, 37].

Kernel PCA

Figure 2.4: Basic idea of kernel PCA.

As illustrated in Fig. 2.4, the idea of Kernel PCA (KPCA) [34] is to nonlinearly map the data $x_1, \dots, x_l \in \mathbb{R}^n$ from the input space to the high dimensional feature space $\Phi(x_1), \dots, \Phi(x_l) \in \mathbb{R}^m$ and then PCA is performed in this feature space.

The mapping can be made by using implicit kernel functions which represent the dot products of vectors in the feature space, i.e.

$$K_{ij} = k(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (2.10)$$

Assume that the data in the feature space are centered, i.e. $\sum_{i=1}^l \Phi(x_i) = 0$, the covariance matrix for the mapped data is defined as

$$C = \frac{1}{l} \sum_{i=1}^l \Phi(x_i) \cdot \Phi(x_i)^T \quad (2.11)$$

To implement PCA, the eigenvalues $\lambda \leq 0$ and eigenvectors $\mathbf{V}$ must be solved from the equation of $\lambda \mathbf{V} = C \mathbf{V}$. It can be proven that there exist coefficients $\alpha_1, \dots, \alpha_l$ such that $\mathbf{V} = \sum_{i=1}^l \alpha_i \Phi(x_i)$. PCA can be applied in the feature space without knowing the explicit form of feature vectors because the PCA projection in the feature space can be obtained by solving the eigenvalue problem of the following equation

$$l \lambda \alpha = K \alpha \quad (2.12)$$

where K is an $l \times l$ matrix with elements K_{ij} and α denotes the column vectors with entries $\alpha_1, \dots, \alpha_l$. The PCA projection of a test sample $\Phi(x)$ onto the k th eigenvector $\mathbf{V}^k$ can be obtained by

$$(\mathbf{V}^k \cdot \Phi(x)) = \sum_{i=1}^l \alpha_i^k (\Phi(x_i) \cdot \Phi(x)) \quad (2.13)$$

By selecting proper kernel functions, we can indirectly obtain desirable nonlinear mappings. It has been proven that any function satisfying the Mercer's theorem can be used as a kernel [38]. There are a few kernels which have been successfully used in Support Vector Machines (SVMs) such as polynomial kernels

$$k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^d \quad (2.14)$$

radial basis kernels

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right) \quad (2.15)$$

sigmoid kernels (with gain κ and offset Θ)

$$k(\mathbf{x}, \mathbf{y}) = \tanh(\kappa(\mathbf{x} \cdot \mathbf{y}) + \Theta) \quad (2.16)$$

In [15, 35], the polynomial kernels are used to take into account higher-order statistics. Experimental results on benchmark databases show that face recognition based on ICA outperforms the conventional PCA approach.

Kernel Discriminant Analysis

Similarly, nonlinear discriminant analysis can also be achieved by using kernels. In [36], the Generalized Discriminant Analysis (GDA) was proposed to generalize LDA to nonlinear problems. The main idea of GDA is to nonlinearly map the data from the input space to the feature space with linear properties in which the classical LDA can be implemented. It has been proven that nonlinear mappings in discriminant analysis can be obtained through kernel operators. Nonlinear discriminant analysis (or kernel Fisherfaces) has been employed in face recognition [37]. To deal with Small Sample Size (SSS) problem in face recognition, the Kernel Direct Discriminant Analysis (KDDA) was proposed in [39]. Their experimental results on the multiview UMIST face database show that the KDDA outperforms the KPCA and GDA approaches.

2.1.5 Edge Map

Cognitive psychological studies show that human beings can recognize line drawings as quickly as gray-level images. Therefore, some researchers argue that edge images can be used for recognition and similar accuracy can be achieved as gray-level images. Edge map is a binary template matching method which employs

holistic similarity measures based on edge information to extract shape and geometrical facial features [40]. Here, we simply regard it as an appearance-based one since it is a holistic template matching method. However, it is mainly based on facial shape information instead of high dimensional space analysis of gray-level images. In [40], face images are first preprocessed to binary images by using Sobel edge detection. The “doubly” modified Hausdorff distance (M2HD), which is less sensitive to small, non-rigid local distortions, was proposed to measure the similarity of two point sets. Given two finite point sets $A = \{a_1, \dots, a_p\}$, and $B = \{b_1, \dots, b_q\}$, the Hausdorff distance is defined as

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (2.17)$$

where

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (2.18)$$

The “doubly” modified Hausdorff distance (M2HD) is defined as

$$d(a, B) = \max(I \min_{b \in N_B^a} \|a - b\|, (1 - I)P) \quad (2.19)$$

$$h(A, B) = \frac{1}{N_a} \sum_{a \in A} d(a, B) \quad (2.20)$$

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (2.21)$$

where N_B^a is a neighborhood of point a in set B , P is associated penalties and N_a is the number of points in set A . I is an indicator $I = 1$ if there exists a point $b \in N_B^a$, and $I = 0$ otherwise.

An upgraded edge map approach, Line Edge Map (LEM), was proposed in [41]. The LEM approach integrates structural information with spatial information by grouping pixels of the face edge map to line segments. A polygonal line fitting process is applied to generate the LEM of a face after thinning the edge map. The

LEM representation only records the end points of line segments on curves such that the storage requirement is further reduced (less than 30% storage space of face edge map). Similarity measures are based on the Line Segment Hausdorff Distance (LHD) which incorporates spatial information, structural information of line orientation and line-point association, and number of disparity (i.e., the average ratio of two LEMS, the number of lines located outside the similarity neighborhood to the total line number). Experimental results on AR face database and Bern University face database show that LEM performs equally well as or superior to the eigenface approach. One of the advantages of the edge map approach is that it is less sensitive to illumination variations. Moreover, proper matching algorithms based on spatial and shape information could improve the tolerance to non-rigid distortions. Nevertheless, to our knowledge, a lot of important facial features exist in gray-level information. Therefore, recognition performance will be improved if gray-level features are efficiently employed.

2.1.6 Hidden Markov Models

Hidden Markov Models (HMMs) have been successfully employed in many research areas such as speech recognition, protein/DNA sequence analysis and robot control etc. More recently, HMMs have also been applied in face recognition [42, 43, 44, 45]. HMMs are powerful statistical models used to characterize the statistical properties of a signal. An HMM consists of two interrelated processes: (1) An underlying, unobservable Markov chain having a finite number of states, a state transition probability matrix and an initial state probability distribution; (2) A set of probability density functions associated with each state. Therefore, an HMM is characterized by three parameters,

$$\lambda = (\mathbf{A}, \mathbf{B}, \mathbf{\Pi}) \quad (2.22)$$

where $\mathbf{A}$ is the state transition probability matrix, $\mathbf{B}$ is the observation probability matrix and $\mathbf{\Pi}$ is the initial state distribution matrix. In practice, the joint likeli-

hood $P(\mathbf{O}, Q|\lambda)$ of a state sequence Q and the corresponding observation sequence $\mathbf{O}$ cannot be evaluated because the state sequence is unknown (i.e. hidden). However, the likelihood $P(\mathbf{O}|\lambda)$ can be evaluated by summing over all possible state sequences. The key attraction of HMMs is that there is a simple procedure for finding the parameters λ which maximize $P(\mathbf{O}|\lambda)$ which is called Baum-Welch re-estimation [46]. For recognition, it is more convenient to maximize $P(\mathbf{O}, Q|\lambda)$ and use the corresponding maximum likelihood. This can be done by using the Viterbi algorithm, a simple dynamic programming optimization procedure [47]. Generally speaking, HMMs are used for coherent 1D signals, while face images are represented by 2D matrices. It will be computationally very complex if they are extended to fully connected two dimensional HMMs. In [42, 43], each face image is divided into overlapping regions with the same width as the image. These significant facial regions (hair, forehead, eyes, nose, mouth, etc) come in a natural order from top to bottom. Each of these facial regions is assigned to a state in a 1D continuous HMM (See Fig. 2.5). In [43], the 2D-DCT coefficients are used to compose observation feature vectors to reduce the computational complexity. Pseudo 2-D HMMs have been also used for face recognition [42, 44, 45]. Different from the 1D HMMs, a sampling window is used to scan the face images top to bottom and left to right. Adjacent sampling windows are also overlapped to improve the ability of the HMM to model the neighborhood relations between sampling blocks. In [44, 45], the DCT coefficients instead of pixel values are used as elements of observation vectors to achieve computation efficiency. For HMMs, although face images are divided into regions, face normalization process is still necessary to achieve better performance. Here, the HMM is regarded as an appearance-based approach because there are no local facial feature detection and matching procedures.

Figure 2.5: Top-bottom 5 state HMM [42].

2.2 Feature-based Face Recognition

Compared with appearance-based approaches, feature-based approaches are more difficult to implement and are computationally complex. Nevertheless, feature-based approaches possess more robustness and generalization ability. Feature-based recognition approaches normally attempt to explore geometrical relationships and shape information of important facial features. Since a human face is a non-rigid object and a lot of variations exist, local feature information is very essential for precise recognition. Feature-based approaches usually rely on the detection and matching of local feature points to obtain the useful local facial features and relationships between faces with variations. Some of the early feature-based approaches extract shape and geometrical features simply based on edge information of face images [48]. However, in [49], it has been experimentally demonstrated that gray-level template matching outperforms such feature-based approaches. In our view, gray-level information is very important to improve the recognition performance. As a result, most of current feature-based approaches also combine gray-level information for feature extraction.

2.2.1 Dynamic Links

Figure 2.6: The shape of a wavelet (The real part and the imaginary part) [50].

In [50], the Dynamic Link Architecture (DLA) was proposed for distortion-invariant object recognition. The DLA can be regarded as an extension of classical artificial neural networks. The DLA approach can cope with complex problems well, especially when flexibility is required, because it exploits correlations in the fine-scale temporal structure of cellular signals to group neurons dynamically into higher-order entities. A concrete implementation of DLA is presented in [50]. This approach uses elastic graph matching to find the closest stored model graph according to a specific cost function. Objects are represented with sparse graphs whose vertices are labelled with a multiresolution description in terms of a local power spectrum, and whose edges are labelled with geometrical distances. The vertex labels are represented by the magnitudes of the Gabor-based wavelets (See Fig. 2.6) with five logarithmically spaced frequency levels and eight orientations which are referred to as a “jet” [51]:

$$\mathcal{J}_{\nu\mu}(\vec{x}_0) := |(\mathcal{W}\mathcal{I}(\vec{k}_{\nu\mu}, \vec{x}_0)| \quad (2.23)$$

where

$$(\mathcal{W}\mathcal{I})(\vec{k}, \vec{x}_0) := \int \psi_{\vec{k}}(\vec{x}_0 - \vec{x}) \mathcal{I}(\vec{x}) d^2x \quad (2.24)$$

$$\psi_{\vec{k}}(\vec{x}) = \frac{\vec{k}^2}{\sigma^2} \exp\left(-\frac{\vec{k}^2 \vec{x}^2}{2\sigma^2}\right) [\exp(i\vec{k}\vec{x}) - \exp(-\frac{\sigma^2}{2})] \quad (2.25)$$

and $\nu \in \{0, \dots, 4\}$ and $\mu \in \{0, \dots, 7\}$. The similarity of jets $\mathcal{J}^I$ in image domain and $\mathcal{J}^M$ in model domain is defined as

$$\mathcal{S}_v(\mathcal{J}^I, \mathcal{J}^M) := \frac{\mathcal{J}^I \cdot \mathcal{J}^M}{\|\mathcal{J}^I\| \|\mathcal{J}^M\|} \quad (2.26)$$

The similarity of edge labels is defined as

$$\mathcal{S}_e(\vec{\Delta}_{ij}^I, \vec{\Delta}_{ij}^M) := (\vec{\Delta}_{ij}^I - \vec{\Delta}_{ij}^M)^2 \quad (2.27)$$

where $\vec{\Delta}_{ij} := \vec{x}_j - \vec{x}_i$ are labelled edges between vertices $\vec{x}_i$ and $\vec{x}_j$ with the Euclidean distance vector. The elastic matching is to search for a set $\{x_i^I\}$ of vertex positions which simultaneously optimize the matching of vertex labels and of edge labels. The total cost function is defined as

$$\mathcal{C}_{total}(\{x_i^I\}) := \lambda \sum \mathcal{S}_e(\vec{\Delta}_{ij}^I, \vec{\Delta}_{ij}^M) - \sum \mathcal{S}_v(\mathcal{J}^I, \mathcal{J}^M) \quad (2.28)$$

where the coefficient λ controls the rigidity of the image graph. The matching is based on a two-stage optimization process (global move and local distortion), varying the image graph to minimize the cost function $\mathcal{C}_{total}$. Fig. 2.7 shows an example of matching results.

Figure 2.7: Example of matching [50].

The Dynamic Link Matching (DLM) for face recognition was proposed in [52]. Similarly, it consists of an image domain and a model domain. Both domains have the form of neural sheets of hypercolumns in which the Gabor-based wavelets are

also used as feature detectors. The image domain and the model domain are bi-directionally connected by dynamic links. DLM serves as a process to restructure the connectivity matrices and to find the correct mapping between the model and the image. The match of the image to the model is performed by network self-organization. A simple winner-take-all mechanism is used to sequentially suppress the least competitive model, and eventually only the best model survives. Once DLM has found the correct mapping between the model and the image, recognition can easily be performed by comparing jets of connected neurons. The limitation of this approach is its computational cost.

2.2.2 Elastic Bunch Graph Matching

The Elastic Bunch Graph Matching (EBGM) was proposed in [53]. Similar to DLA, objects are represented by labelled graphs: edges are labelled with distance information and nodes are labelled with wavelet responses locally bundled in jets. The nodes are manually positioned at a set of fiducial points, e.g. the pupils, the corners of the mouth, the tip of the nose, the top and bottom of the ears etc (See Fig. 2.8). Different from [50] in which only the magnitude information is used, the phase information is also used in the similarity function of jets to achieve accurate localization. The similarity function is given by

$$\mathcal{S}_\phi(\mathcal{J}, \mathcal{J}') = \frac{\sum_j a_j a'_j \cos(\phi_j - \phi'_j - \vec{d} \vec{k}_j)}{\sqrt{\sum_j a_j^2 \sum_j a'^2_j}} \quad (2.29)$$

where a_j and ϕ_j is the j th magnitude and phase of the jet $\mathcal{J}$ respectively and $\vec{d}$ is the relative displacement of jets $\mathcal{J}$ and $\mathcal{J}'$. In order to find fiducial points in new faces, a representative set of individual model graphs are combined into a stack-like structure, called a Face Bunch Graph (FBG) (See Fig. 2.9). A set of jets referring to one fiducial point is called a bunch. The goal of Elastic Bunch Graph Matching on a probe image is to find the fiducial points and thus to extract from the image a graph which maximizes the similarity with the FBG. The graph

similarity between an image graph and the FBG of identical pose depends on the jet similarity and distortion of the image grid relative to the FBG grid. Therefore, the similarity of image graph $\mathcal{G}^I$ and an FBG $\mathcal{B}$ is defined as

$$\mathcal{S}_B(\mathcal{G}^I, \mathcal{B}) = \frac{1}{N} \sum_n \max_m (\mathcal{S}_\phi(\mathcal{J}_n^I, \mathcal{J}_n^{B_m})) - \frac{\lambda}{E} \sum_e \frac{(\Delta \vec{x}_e^I - \Delta \vec{x}_e^B)^2}{(\Delta \vec{x}_e^B)^2} \quad (2.30)$$

where λ determines the relative importance of jets and metric structure, and $\Delta \vec{x}_e$ are the distance vectors used as labels at edges e . The matching procedure is based on a coarse to fine approach including four steps: (1) Find approximate face position; (2) Refine position and size; (3) Refine size and find aspect ratio; (4) Implement local distortion. After obtaining the model graphs from the gallery images and the image graphs from the probe images, recognition can be carried out. The recognition is based on the graph similarity of image graph $\mathcal{G}^I$ and model graph $\mathcal{G}^M$:

$$\mathcal{S}_G(\mathcal{G}^I, \mathcal{G}^M) = \frac{1}{N'} \sum_{n'} \mathcal{S}_a(\mathcal{J}_{n'}^I, \mathcal{J}_{n'}^M) \quad (2.31)$$

For recognition, the jet similarity function without phase is used because it is more robust against changes in facial expression and other variations. The EBGM is a successful approach which can be used to deal with a lot of variations in face recognition, e.g. translation, rotation, scale and deformation, etc. Therefore, compared with the classic appearance-based approach, it achieves more generalization such that only a small number of examples are required to handle the complex task of face recognition.

2.2.3 Active Appearance Model

A compact parametrized model of facial appearance with both shape and gray-level information is presented in [54]. First, the shapes of facial features and their spatial relationships are represented by a single flexible shape model, i.e. Point Distribution Model (PDM). In each face image, the main facial features are

Figure 2.8: Labelled graph with different poses [53].

Figure 2.9: Face bunch graph (FBG) [53].

delineated by a large number of landmark points (As shown in Fig. 2.10). The flexible shape model is statistically generated from a set of training samples. Based on the Principal Component Analysis (PCA), the shape model is denoted as

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b} \quad (2.32)$$

where $\mathbf{x} = (x_1, y_1, \dots, x_n, y_n)^T$ is a vector consisting of positions of n landmark points and (x_k, y_k) is the coordinate of the k th landmark point. $\bar{\mathbf{x}}$ represents the mean shape. $\mathbf{P}$ is a matrix of eigenvectors and $\mathbf{b}$ is a vector of eigenvector weights. By modifying $\mathbf{b}$, new instances of the model can be generated. Furthermore, in

order to let the gray-level appearance be independent of shape (shape-free), each gray-level face image is deformed into a mean shape such that changes in gray-level intensities are kept to a minimum. The training of the deformed gray-level images is also based on PCA (eigenface) (2.32). The gray-level intensity values instead of coordinates of landmark points are used. In order to locate facial features in the new face image, local gray-level profile models are also built to represent local facial features. The shape model and local gray-level models are both used to automatically locate all the modelled features simultaneously. This is achieved using an Active Shape Model (ASM) search [55]. In the fitting procedure, the pose parameters (translation, scaling, and rotation) are first adjusted to minimize the mean squared distance between model points and the suggested new positions. Then the shape model is deformed to fit the suggested points based on the least squares approximation by modifying the shape parameters $\mathbf{b}$. After the shape model is fitted to a new face image presented, the appearance parameters, i.e. shape, local gray-level and shape-free gray-level, can be computed. Face recognition can be performed using the appearance parameters individually, or in combination.

Figure 2.10: A labelled training image gives a shape free patch and a set of points [56].

However, the above approach is not guaranteed to give an optimal fit of the appearance model to the image because small errors in the match of the shape model may result in a shape-normalized texture map that cannot be reconstructed correctly. The Active Appearance Model (AAM) was proposed in [56, 57, 58]. This approach is an efficient direct optimization approach which matches shape and

Figure 2.11: Multiresolution search from displaced position using face model [56].

texture (intensities or colors) simultaneously. Similar to the approach of [54], a shape model and a texture model (called gray-level model in [54]) are built based on PCA and then combined into an appearance model. The model is given by

$$\begin{aligned} \mathbf{x} &= \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c} \\ \mathbf{g} &= \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c} \end{aligned} \quad (2.33)$$

where $\mathbf{c}$ are parameters controlling the shape and texture of the appearance model. The matching is based on an iterative process to optimize the difference between the given image and the image generated by the model (called residual), i.e.

$$\mathbf{r}(\mathbf{p}) = \mathbf{g}_s - \mathbf{g}_m \quad (2.34)$$

where $\mathbf{g}_s = T_{\mathbf{u}}^{-1}(\mathbf{g}_{im})$ is the image sample at current estimate $\mathbf{g}_{im}$ projected into the texture model frame, $\mathbf{g}_m = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c}$ is the current model texture and $\mathbf{p}$ are the parameters of the model including model parameters $\mathbf{c}$, the pose $\mathbf{t}$ and the texture transformation $\mathbf{u}$. The optimization procedure is repeated until no improvement is made with respect to the error $\|\mathbf{r}\|^2$. In practice, a multiresolution implementation is used to improve the speed and robustness. Fig. 2.11 shows frames from an

AAM search for two faces. The approaches discussed here can also be regarded as model-based approaches. We regard them as feature-based approaches because shape features play a key role in these approaches and normally main facial features are needed to build the models. Not like early-stage feature-based approaches in which only position information between facial features is used, these approaches also combine gray-level or texture information to achieve more accurate recognition performance.

2.3 Neural Networks

Neural networks have been widely applied in face recognition. The aforementioned DLA is a special example of neural networks application in face recognition. In [59], the authors proposed a hybrid neural network solution for face recognition, which combines local image sampling, a self-organizing map (SOM) neural network, and a convolutional neural network [60]. The self-organizing map neural network is for dimensionality reduction and the convolutional neural network is used to extract features in a hierarchical set of layers and is incorporated with a multi-layer perceptron (MLP) style classifier for classification. In [61], a face recognition/detection system based on Probabilistic Decision-Based Neural Networks (PDBNN) was proposed. It adopts a hierarchical network structure with nonlinear basis functions and a competitive credit-assignment scheme. PDBNN-based recognition systems possess the merits of neural networks and statistical approaches.

Neural networks can be used for both feature extraction and classification. For classification, they have been proven to possess a lot of advantages because of their learning ability and good generalization [9, 59, 60, 61, 62, 63, 64]. They are able to classify patterns well even when the distribution of patterns in the multi-dimensional space is very complex. Compared with the basic Nearest Neighbor (NN) or k -Nearest Neighbor (KNN) classifiers, neural-network-based approaches

are more efficient in terms of memory space and computation time and they are more fault tolerant and insensitive to irrelevant and noisy parameters. Multi-Layer Perceptrons (MLPs) coupled with the back-propagation (BP) algorithm are widely used in face recognition [63]. The main limitations of MLP are attributed to the slow convergence speed and local minima. In [65], Radial Basis Function (RBF) neural networks have been successfully employed in face recognition. The main advantages of RBF neural networks are their compact topology and locally tuned neurons such that fast learning speed can be achieved. In order to overcome the limitations of MLPs, the authors employ a hybrid learning algorithm to train the RBF neural networks such that the dimension of the search space is drastically reduced in the gradient paradigm and it achieves high recognition rate and good learning efficiency. It should be noted that because face images are high dimensional data, dimensionality reduction before classification based on neural networks is essential to avoid the overfitting problem.

Recently, Support Vector Machines (SVMs), another category of universal feed-forward networks proposed by Vapnik [66], has attracted a lot of attention for pattern classification. It has also been successfully employed in face detection and recognition applications [67, 68, 69]. SVMs are primarily two-class classifiers which aim to find the optimal separating hyperplane such that the expected generalization error is minimized. Most methods for training a classifier by minimizing the empirical risk are based on training data. Instead, SVMs perform structural risk minimization to achieve good generalization (i.e. minimize an upper bound on the expected generalization error). For face recognition, SVMs should be extended to multi-class classification. There are two schemes for solving multi-class problems with SVMs. In [68], the one-against-one strategy is employed while the one-against-all strategy is used in [69]. In many applications including face recognition, the data are non-linearly separable. SVMs can be extended to non-linear classifiers by using kernel-based methods similar to the KPCA described in Section 2.1.4. Polynomial, RBF and Sigmoid kernels have been successfully employed in SVMs (2.14), (2.15), (2.16). The major limitations of SVMs lie in the choice of

the kernel, speed and size especially in training.

Higher-order neural networks (HONN) have been applied in image recognition [70, 71, 72, 73, 74, 75, 76]. The most important advantage of HONN is that invariances can be built directly into the architecture of HONN such that it is invariant to geometric transformations, such as translation, scale, or in-plane rotation. It has been shown that a third-order network can be used to achieve such invariance. However, the most severe limitation of third-order networks is that the number of triplet combinations dramatically increases as the size of the input field increases. To circumvent this limitation, in [75] the coarse coding method based on binary images was proposed to represent the high resolution images using a set of low resolution images such that the input fields for each image are largely reduced. In [77], the third-order neural networks were employed for face recognition. Since HONN is usually for binary images, the gray level face images are represented using a set of quantized isodensity binary images and thus the coarse coding algorithm can be employed for the isodensity images to reduce the memory requirements and computational complexity. The major drawback in using isodensity information is that it is very sensitive to lighting variations on face images although the performance can be improved using appropriate preprocessing methods such as histogram equalization.

2.4 Face Databases

In order to test face recognition algorithms, many public face databases have been built. Appropriate databases should be chosen based on different evaluation tasks such as pose, facial expression, illumination and occlusion, etc. To facilitate the development of subsequent chapters, a few well-known face databases are summarized in Table 2.1.

Table 2.1: Face Databases

Face database	No. of subjects	No. of images	Original image resolution	Brief description
ORL [78]	40	400	92×112	Grayscale images; poses, illumination, expressions, w/ and w/o glasses.
Yale [18]	15	165	320×243	Grayscale images; expressions, illumination, w/ and w/o glasses.
Yale B [79]	10	5760	640×480	Grayscale images; poses, illumination (lighting /pose angles provided).
CMU PIE [80]	68	41368	640×486	Color images; illumination, poses, expressions (lighting /pose angles provided).
FERET [4, 5]	1199	14051	256×384	Grayscale images; poses, expressions, illumination, aging.
AR [81]	126	4000	768×576	Color images; illumination, expressions, hair style, occlusions (glasses/scarf).
UMIST [82]	20	564	220×220	Grayscale images; poses.
XM2VTS [83]	295	2360	720×576	Color images; speech, 3D models, video sequences (speaking/rotating head).
MIT [13]	16	433	120×128	Grayscale images; scale, head rotation, illumination.

Chapter 3

PCA and LDA in Discrete Cosine Transform Domain

In this chapter, we prove that Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) can be directly implemented in the Discrete Cosine Transform (DCT) domain and the results are exactly the same as the one obtained from the spatial domain. In some applications, compressed images are desirable to reduce the storage requirement. For images compressed using the DCT, PCA and LDA can be directly implemented in the DCT domain such that the inverse DCT transform can be skipped and the dimensionality of the original data can be initially reduced to cut down computational cost.

3.1 Introduction

Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) have been widely used for dimensionality reduction or feature extraction in pattern recognition [84]. They are also successfully employed in face recognition research area, i.e., Eigenfaces and Fisherfaces [13, 18]. Dimensionality reduction is essential for extracting effective features and reducing computational complexity

in classification stage. However, dimensionality reduction using PCA and LDA is also computationally expensive when the original dimensionality is high and the number of training samples is large. When the training database becomes larger, the training time and memory requirement will significantly increase. Moreover, systems based on PCA or LDA should be retrained when new classes are added to obtain optimal projection results. Therefore, reduction in computational complexity is highly desirable. More recently, Discrete Cosine Transform (DCT) has been employed in face recognition for dimensionality reduction [85, 86]. The advantage of the DCT is that it is data independent (i.e., the basis images are only dependent on one image instead of on the entire set of training images) and it can be implemented using a fast algorithm. Nevertheless, only limited low-frequency coefficients are used as features if the DCT is employed for direct dimensionality reduction. Although the DCT is asymptotically equivalent to the Karhunen-Loeve Transform (KLT) (PCA) for Markov-1 signals with a correlation coefficient that is close to one [87], PCA is more efficient for dimensionality reduction since it is the optimal dimensionality reduction method for Gaussian distributed data. In [88], the authors have experimentally shown that PCA can be applied on reduced DCT coefficients to achieve better recognition performance.

In this chapter, we demonstrate that PCA and LDA can be directly implemented in the DCT domain via theory and the results are exactly the same as the one obtained from the spatial domain. In some applications, the DCT can be first employed to remove redundant information and PCA or LDA can be subsequently implemented in the DCT domain such that the computational complexity can be significantly reduced. Furthermore, PCA or LDA may be directly implemented in the block DCT domain which is widely used in the JPEG and MPEG standard. There are two advantages if PCA or LDA can be implemented in the block DCT domain: 1) If the database is stored as compressed JPEG images to reduce storage requirement, DCT coefficients of JPEG images can be directly used such that the inverse DCT can be skipped to cut down computational cost; 2) In JPEG images, some redundant information has been removed by quantization such that

the dimensionality of images can be initially reduced by removing coefficients with less information.

The rest of this chapter is organized as follows. In Section 3.2, It is proven that PCA can be implemented in the DCT or the block DCT domain. Similarly, it is proven that LDA and D-LDA can also be implemented in the DCT or the block DCT domain in Section 3.3 and Section 3.4 respectively. Experimental results on the FERET database are shown in Section 3.5. Finally, conclusions are given in Section 3.6.

3.2 PCA in DCT Domain

3.2.1 PCA on Orthonormally Transformed Data

In the following section, we show that the PCA projection result is invariant for the orthonormally transformed data. Combining with the subsequent proof that the DCT is an orthogonal transformation, PCA can be directly implemented in the DCT domain.

Theorem 1 *The PCA subspace projection result remains the same if the original data are transformed using an orthonormal transformation.*

Proof. Let $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$ denote an n -dimensional random vector. The mean of the random vector $\mathbf{x}$ is denoted by $\bar{\mathbf{x}} = E[\mathbf{x}]$ and the covariance matrix of the random vector $\mathbf{x}$ is defined as $C_{\mathbf{x}} = E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T]$. The PCA projection matrix A can be obtained by eigen-analysis of the covariance matrix $C_{\mathbf{x}}$, i.e.

$$C_{\mathbf{x}}a_i = \lambda_i a_i, \quad i = 1, 2, \dots, m \quad (3.1)$$

where a_i is the i th largest eigenvector of $C_{\mathbf{x}}$, $m < n$ and $A = [a_1, a_2, \dots, a_m] \in \mathbb{R}^{n \times m}$. We have the following projection result:

$$\mathbf{y} = A^T(\mathbf{x} - \bar{\mathbf{x}}) \quad (3.2)$$

Let us assume that the original data are transformed by using an $n \times n$ orthogonal matrix Q , i.e., $\mathbf{z} = Q^T \mathbf{x}$. The mean of the transformed random vector $\mathbf{x}$ is given by

$$\begin{aligned}\bar{\mathbf{z}} &= E[Q^T \mathbf{x}] \\ &= Q^T E[\mathbf{x}] \\ &= Q^T \bar{\mathbf{x}}\end{aligned}\tag{3.3}$$

We may obtain the covariance matrix $C_{\mathbf{z}}$ of the transformed random vector $\mathbf{z}$ as follows:

$$\begin{aligned}C_{\mathbf{z}} &= E[(\mathbf{z} - \bar{\mathbf{z}})(\mathbf{z} - \bar{\mathbf{z}})^T] \\ &= E[(Q^T \mathbf{x} - Q^T \bar{\mathbf{x}})(Q^T \mathbf{x} - Q^T \bar{\mathbf{x}})^T] \\ &= Q^T E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T] Q \\ &= Q^T C_{\mathbf{x}} Q\end{aligned}\tag{3.4}$$

Similarly, the PCA projection matrix $\tilde{A} = [\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_m] \in \mathbb{R}^{n \times m}$ for the orthonormally transformed data can be obtained by eigen-analysis of the covariance matrix $C_{\mathbf{z}}$ and we have

$$C_{\mathbf{z}} \tilde{a}_i = \lambda_i \tilde{a}_i, \quad i = 1, 2, \dots, m\tag{3.5}$$

Substituting (3.4) into (3.5), we have

$$Q^T C_{\mathbf{x}} Q \tilde{a}_i = \lambda_i \tilde{a}_i\tag{3.6}$$

Since Q is an orthogonal matrix, i.e., $Q^T = Q^{-1}$, equation (3.6) can be rewritten as

$$C_{\mathbf{x}} Q \tilde{a}_i = \lambda_i Q \tilde{a}_i\tag{3.7}$$

Equation (3.7) shows that the covariance matrices $C_{\mathbf{x}}$ and $C_{\mathbf{z}}$ have exactly same eigenvalues and the relation of their eigenvectors is $a_i = Q\bar{a}_i$. Therefore, the PCA projection matrices A and $\bar{A}$ also satisfy $\bar{A} = Q^T A$. The projection result of $\mathbf{z}$ is as follows:

$$\begin{aligned}\bar{\mathbf{y}} &= \bar{A}^T(\mathbf{z} - \bar{\mathbf{z}}) \\ &= A^T Q(\mathbf{z} - \bar{\mathbf{z}}) \\ &= A^T(\mathbf{x} - \bar{\mathbf{x}}) = \mathbf{y}\end{aligned}\tag{3.8}$$

From (3.8), we can conclude that for the PCA subspace projection, orthonormal transformation of the original data will not change the projection result.

3.2.2 Discrete Cosine Transform

There are four established types of Discrete Cosine Transforms (DCTs), i.e., DCT-I, DCT-II, DCT-III and DCT-IV [87]. The DCT-II is more widely applied in signal coding. Hence, the DCT-II is often simply referred to as “the DCT”. In this chapter, only the DCT-II is discussed.

In the 1D case, the DCT is defined as

$$y(k) = \alpha(k) \sum_{n=0}^{N-1} x(n) \cos\left[\frac{\pi(2n+1)k}{2N}\right], \quad k = 0, 1, \dots, N-1 \tag{3.9}$$

where

$$\alpha(k) = \begin{cases} \frac{1}{\sqrt{N}}, & k = 0 \\ \sqrt{\frac{2}{N}}, & k = 1, 2, \dots, N-1 \end{cases}$$

The DCT can be written in a vector form as $\mathbf{y} = C^T \mathbf{x}$. The elements of the matrix

C are

$$\begin{aligned} c_{n,k} &= \frac{1}{\sqrt{N}}, \quad k = 0, \quad n = 0, 1, \dots, N-1 \\ c_{n,k} &= \sqrt{\frac{2}{N}} \cos\left[\frac{\pi(2n+1)k}{2N}\right], \quad k = 1, 2, \dots, N-1, \quad n = 0, 1, \dots, N-1 \end{aligned}$$

It has been proven that the transform matrix C is an orthogonal matrix [87], i.e., $C^T = C^{-1}$. The $M \times N$ 2D DCT and inverse DCT (IDCT) may be computed using the separable 1D row and column transformations, i.e.

$$Y = C_M^T X C_N \quad (3.10)$$

$$X = C_M Y C_N^T \quad (3.11)$$

For the PCA applied for images, a 2D image matrix is normally converted into a vector by concatenating the columns or rows of the matrix. Accordingly, we may convert the 2D DCT and IDCT shown in (3.10) and (3.11) into the following vector form:

$$\hat{Y} = G^T \hat{X} \quad (3.12)$$

$$\hat{X} = H^T \hat{Y} \quad (3.13)$$

where $\hat{Y}$ and $\hat{X}$ are MN -dimensional vectors, and G and H are $MN \times MN$ transformation matrices. If $\hat{X}$ is obtained by concatenating the columns of X , i.e.

$$\hat{X} = [x_{0,0} \dots x_{M-1,0} \dots x_{M-1,N-1}]^T \quad (3.14)$$

$$\hat{Y} = [y_{0,0} \dots y_{M-1,0} \dots y_{M-1,N-1}]^T \quad (3.15)$$

it can be derived from (3.10) that the transformation matrix is as follows:

$$G = \begin{bmatrix} c_{0,0}c_{0,0} & \cdots & c_{0,M-1}c_{0,0} & \cdots & \cdots & c_{0,M-1}c_{0,N-1} \\ \vdots & & \vdots & & & \vdots \\ c_{M-1,0}c_{0,0} & \cdots & c_{M-1,M-1}c_{0,0} & \cdots & \cdots & c_{M-1,M-1}c_{0,N-1} \\ \vdots & & \vdots & & & \vdots \\ \vdots & & \vdots & & & \vdots \\ c_{M-1,0}c_{N-1,0} & \cdots & c_{M-1,M-1}c_{N-1,0} & \cdots & \cdots & c_{M-1,M-1}c_{N-1,N-1} \end{bmatrix} \quad (3.16)$$

It can also be derived from (3.10) that the inverse transformation matrix H is exactly the transpose of G , i.e., $G = H^T$. Combining with (3.12) and (3.13), we have

$$G^{-1} = G^T \quad (3.17)$$

Equation (3.17) shows that the transformation matrix G is orthogonal. Furthermore, the sequence of elements in $\hat{Y}$ only corresponds to the sequence of column vectors in G . When the elements in $\hat{Y}$ are permuted, matrix G remains orthogonal.

From Theorem 1 and (3.17), we can conclude that the projection result of PCA in the 2D DCT domain is exactly the same as the one in the 2D spatial domain.

3.2.3 Block DCT

In JPEG standard, images are first divided into rectangular blocks (8×8) [89]. The DCT is applied independently on each block. Given a $pn \times qn$ image which is divided into $n \times n$ subimages, the DCT is applied independently on each subimage. The transformation may be represented as follows:

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{p1} \\ \vdots \\ \vdots \\ Y_{pq} \end{bmatrix} = \begin{bmatrix} G_{11} & & & \\ & \ddots & & \\ & & G_{p1} & \\ & & & \ddots \\ & & & & G_{pq} \end{bmatrix} \begin{bmatrix} X_{11} \\ \vdots \\ X_{p1} \\ \vdots \\ \vdots \\ X_{pq} \end{bmatrix} \quad (3.18)$$

Since the square matrices G_{ij} are all orthogonal, i.e., $G_{ij}^T = G_{ij}^{-1}$, we have

$$\begin{bmatrix} G_{11} & & & \\ & \ddots & & \\ & & G_{p1} & \\ & & & \ddots \\ & & & & G_{pq} \end{bmatrix}^T = \begin{bmatrix} G_{11} & & & \\ & \ddots & & \\ & & G_{p1} & \\ & & & \ddots \\ & & & & G_{pq} \end{bmatrix}^{-1} \quad (3.19)$$

As shown in (3.19), the block diagonal transformation matrix is also an orthogonal matrix. Therefore, JPEG DCT coefficients can be directly used for the PCA approach. The inverse DCT can be skipped.

3.3 LDA in DCT Domain

Similarly, it can be shown that LDA can be directly implemented in the DCT domain.

Theorem 2 *The LDA subspace projection result remains the same if the original data are transformed using an orthonormal transformation.*

Proof. For LDA, the between-class scatter matrix S_b and the within-class scatter matrix S_w are respectively defined in (2.4) and (2.5). The optimal subspace projection is determined as follows:

$$E_{opt} = \arg \max_E \frac{|E^T S_b E|}{|E^T S_w E|} = [e_1, e_2, \dots, e_m] \quad (3.20)$$

where $[e_1, e_2, \dots, e_m]$ is the set of generalized eigenvectors corresponding to the m ($m \leq c - 1$) largest generalized eigenvalues λ_i , $i = 1, 2, \dots, m$, i.e.

$$S_b e_i = \lambda_i S_w e_i, \quad i = 1, 2, \dots, m \quad (3.21)$$

So, the projection result is

$$\mathbf{y} = E_{opt}^T \mathbf{x} \quad (3.22)$$

Let us assume that the original data is transformed using an orthogonal matrix Q , i.e., $\mathbf{z} = Q^T \mathbf{x}$. Similar to (3.4), the between-class and within-class scatter matrices for $\mathbf{z}$ are obtained as follows:

$$\tilde{S}_b = Q^T S_b Q \quad (3.23)$$

$$\tilde{S}_w = Q^T S_w Q \quad (3.24)$$

and

$$\tilde{E}_{opt} = \arg \max_{\tilde{E}} \frac{|\tilde{E}^T \tilde{S}_b \tilde{E}|}{|\tilde{E}^T \tilde{S}_w \tilde{E}|} = [\tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_m] \quad (3.25)$$

We can obtain $\tilde{E}_{opt}$ by solving the generalized eigenvalues and eigenvectors of $\tilde{S}_b$ and $\tilde{S}_w$ from the following equation:

$$\tilde{S}_b \tilde{e}_i = \lambda_i \tilde{S}_w \tilde{e}_i, \quad i = 1, 2, \dots, m \quad (3.26)$$

Substituting (3.23) and (3.24) into (3.26), we have

$$S_b Q \tilde{e}_i = \lambda_i S_w Q \tilde{e}_i, \quad i = 1, 2, \dots, m \quad (3.27)$$

Comparing (3.21) and (3.27), we can see that the generalized eigenvalues are the same and the generalized eigenvectors satisfy $e_i = Q \tilde{e}_i$, $i = 1, 2, \dots, m$. Since Q is orthogonal, the optimal projection matrix satisfies $\tilde{E}_{opt} = Q^T E_{opt}$. Therefore, the projection result of $\mathbf{z}$ is as follows:

$$\begin{aligned} \tilde{\mathbf{y}} &= \tilde{E}^T \mathbf{z} \\ &= (Q^T E_{opt})^T \mathbf{z} \\ &= E_{opt}^T Q \mathbf{z} \\ &= E_{opt}^T \mathbf{x} = \mathbf{y} \end{aligned} \quad (3.28)$$

From (3.28), we can conclude that for the LDA subspace projection, orthonormal transformation of the original data will not change the projection result.

As shown in Section 3.2, the DCT or the block DCT transformation matrices are orthogonal. As a consequence, the result of LDA in the DCT domain is the same as the one in the spacial domain.

3.4 D-LDA in DCT Domain

For LDA in face recognition, there will be singular problems in scatter matrices due to the high-dimensional data and small number of training samples. As a consequence, PCA + LDA (Fisherface) scheme is commonly used [18]. Since the DCT will not change the PCA projection result, the two-stage approach can also be implemented in the DCT domain and the projection result remains the same. If DCT coefficients can be initially discarded to avoid singular problems as well as keep important high-frequency features, LDA can be directly implemented in the truncated DCT domain. This approach is more computationally efficient than the PCA + LDA scheme.

More recently, some variants of LDA have been proposed to solve the small sample size problem [20, 21]. In fact, some variants of LDA can also be directly implemented in the DCT or block DCT domain. Here, we simply prove that the result of Direct LDA (D-LDA) is invariant for the orthonormally transformed data based on [21].

The key idea of the D-LDA is to discard the null space of S_b rather than the null space of S_w . Please refer to [21] for detailed descriptions of D-LDA. The brief steps are as follows:

$$A^T S_b A = D_b \quad (3.29)$$

$$(A D_b^{-1/2})^T S_b (A D_b^{-1/2}) = I \quad (3.30)$$

Diagonalizing $(AD_b^{-1/2})^T S_w (AD_b^{-1/2})$ by eigen-analysis, we have

$$B^T (AD_b^{-1/2})^T S_w (AD_b^{-1/2}) B = D_w \quad (3.31)$$

The D-LDA transformation matrix is $E = AD_b^{-1/2} B$, where D_b is an $m \times m$ diagonal matrix whose diagonal elements are the largest m eigenvalues of S_b , and A is an $n \times m$ matrix which consists of m eigenvectors corresponding to the m largest eigenvalues in D_b . The term D_w is an $l \times l$ diagonal matrix whose diagonal elements are the smallest l eigenvalues of S_w , and $(AD_b^{-1/2}) B$ consists of the eigenvectors corresponding to the smallest l eigenvalues in D_w , where B is an $m \times l$ matrix ($l \leq m \leq n$).

In Section 3.3, it has been proven that the scatter matrices S_b and S_w of the original data and orthonormally transformed data satisfy the relations given by (3.23), (3.24). Similarly, for the orthonormal data $Q^T \mathbf{x}$, we have

$$\tilde{A}^T \tilde{S}_b \tilde{A} = \tilde{A}^T Q^T S_b Q \tilde{A} = D_b \quad (3.32)$$

$$(\tilde{A} D_b^{-1/2})^T Q^T S_b Q (\tilde{A} D_b^{-1/2}) = I \quad (3.33)$$

Diagonalizing $(\tilde{A} D_b^{-1/2})^T \tilde{S}_w (\tilde{A} D_b^{-1/2})$ by eigen-analysis, we have

$$\begin{aligned} & \tilde{B}^T (\tilde{A} D_b^{-1/2})^T \tilde{S}_w (\tilde{A} D_b^{-1/2}) \tilde{B} \\ &= \tilde{B}^T (\tilde{A} D_b^{-1/2})^T Q^T S_w Q (\tilde{A} D_b^{-1/2}) \tilde{B} \\ &= D_w \end{aligned} \quad (3.34)$$

The D-LDA transformation matrix for the orthonormally transformed data is $\tilde{E} = \tilde{A} D_b^{-1/2} \tilde{B}$. It follows from (3.29) and (3.32), (3.31) and (3.34) that $\tilde{A} = Q^T A$ and $\tilde{B} = B$. Accordingly, the projection result of the D-LDA for the orthonormally transformed data is as follows:

$$\tilde{\mathbf{y}} = \tilde{E}^T \tilde{\mathbf{x}}$$

$$\begin{aligned}
&= (\tilde{A}D_b^{-1/2}\tilde{B})^T\tilde{\mathbf{x}} \\
&= (Q^TAD_b^{1/2}B)^TQ^T\mathbf{x} \\
&= (AD_b^{1/2}B)^T\mathbf{x} = \mathbf{y}
\end{aligned} \tag{3.35}$$

From (3.35), we can conclude that for the D-LDA subspace projection, orthonormal transformation of the original data will not change the projection result.

3.5 Experimental Results

Although the proposed approach is not restricted to face recognition applications, in this section, we present some experimental results based on face recognition. The FERET evaluation methodology a benchmark in which many face images with all kinds of variations are available [4, 5]. Therefore, it is used to evaluate the proposed approach. We use the FERET September 1996 Test in which only frontal face images are employed. There are one gallery set and four probe sets (i.e., Duplicate 1, Duplicate 2, Fafb and Fafc) for different evaluation tasks (See Table 3.1). The original face images are first normalized by using the preprocessing program provided in the CSU evaluation system [90] which is an improved version of [4, 5]. Examples of normalized face images (150×130) are shown in Fig. 3.1. Cumulative match curves based on the Euclidean distance are employed to evaluate the performance. The cumulative match curves are generated by first computing the rank of the closest gallery image of the same subject for each probe image and then summing the number of correct matches for each rank. The “rate” in the following figures is the portion of the correct matches in the total number of images of each probe set. In the following experiments for PCA, 50 principal components are used since they are enough to achieve similar performance as complete principal components used.

As illustrated in Fig. 3.2, the coefficient feature vector can be formed by independently scanning each coefficient block in a zigzag manner and then concatenating

Table 3.1: Four probe subsets and their evaluation task

Probe subsets	Evaluation task
Dup 1	Aging of subjects
Dup 2	Aging of subjects
Fafb	Facial expression
Fafc	Illumination

them. Fortunately, in JPEG standard, DCT coefficients are also coded in a zigzag order because DCT coefficients with great magnitude are mainly located at the upper-left corner of each block. It should be noted that in the experiments, the JPEG DCT coefficients are quantized using a typical quantization matrix shown in Fig. 3.3 [91].

By using PCA or LDA in the DCT domain, not only the storage requirement is largely relaxed, but also the computational complexity is simplified. Some redundant information may be first removed by truncating the DCT coefficients so that the dimensionality of the coefficient vectors can be reduced. For PCA and LDA, the computational complexity, which mainly depends on the dimensionality of the original data, can be largely reduced.

In each block, coefficients with less information can be initially discarded. The number of discarded coefficients can be determined according to the quantization procedure since a lot of coefficients become zero after quantization. However, we cannot simply discard the zero coefficients because their positions vary with different blocks. In our approach, coefficients are retained following the zigzag scan and the number of retained coefficients is the same for each block.

Figs. 3.4 – 3.7 show the performance of the PCA on original uncompressed images and on quantized JPEG DCT coefficient vectors (64 complete coefficients and 20 coefficients of each block are respectively illustrated). Their performances are quite similar. However, for a 128×128 image, the dimension is reduced from 16384 to 5120 if only 20 coefficients for each block are used. For the PCA method, the computational complexity is around $O(n^2 + m^3)$, where n is the input dimension of PCA projection and m is the output dimension of PCA projection normally

Figure 3.1: Examples of normalized FERET face images from four subsets: (a) Dup 1; (b) Dup 2; (c) Fafb; (d) Fafc

in low dimension. Therefore, the computational cost and memory requirements can be cut down by reducing n . The results of the D-LDA method are depicted in Figs. 3.8 – 3.11. The performance of the quantized JPEG coefficients is even slightly better than the performance of the original uncompressed images because some irrelevant information has been initially removed by quantization such that the generalization of D-LDA is improved. Therefore, dimensionality can be first reduced by appropriately discarding DCT coefficients with less information to save the computational cost for PCA and LDA (D-LDA). It is interesting to note that the performance will not be affected and in fact, can even be improved.

Figure 3.2: Feature vector formation in block DCT domain.

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Figure 3.3: A typical quantization matrix.

3.6 Conclusion

This chapter presents an efficient way of reducing storage requirements and computational complexity of PCA and LDA for pattern recognition applications. It has been proven that PCA and LDA can be directly implemented in the DCT domain or block DCT domain. Therefore, JPEG images can be used to reduce storage requirements. By initially discarding an appropriate number of DCT coefficients with less information, the computational complexity of PCA and LDA can be significantly reduced which facilitates the implementation of a real-time pattern recognition system.

Figure 3.4: Performance of PCA in JPEG DCT domain (Dup 1).

Figure 3.5: Performance of PCA in JPEG DCT domain (Dup 2).

Figure 3.6: Performance of PCA in JPEG DCT domain (Fafb).

Figure 3.7: Performance of PCA in JPEG DCT domain (Fafc).

Figure 3.8: Performance of D-LDA in JPEG DCT domain (Dup 1).

Figure 3.9: Performance of D-LDA in JPEG DCT domain (Dup 2).

Figure 3.10: Performance of D-LDA in JPEG DCT domain (Fafb).

Figure 3.11: Performance of D-LDA in JPEG DCT domain (Fafc).

Chapter 4

High-Speed Face Recognition Based on Discrete Cosine Transform and RBF Neural Networks

In this chapter, an efficient method for high-speed face recognition based on Discrete Cosine Transform (DCT), Linear Discriminant Analysis (LDA) and RBF neural networks is presented. First, the dimensionality of the original face image is reduced by using the DCT. Next, the truncated DCT coefficient vectors are clustered using the proposed clustering algorithm. This process makes the subsequent LDA more efficient. After implementing LDA, the most discriminating and invariant facial features are maintained and the training samples are clustered well. As a consequence, further parameter estimation for the RBF neural networks is fulfilled easily which facilitates fast training in the RBF neural networks. Simulation results show that the proposed system achieves excellent performance with high training and recognition speed, high recognition rate as well as good illumination robustness.

4.1 Introduction

As mentioned in Chapter 2, research on face recognition can be grouped into two categories, namely, feature-based and appearance-based approaches. Appearance-based face recognition has attracted more attention since the well-known statistical method, the Principal Component Analysis (PCA) (also known as Karhunen-Loeve Transform (KLT)) was applied in face recognition [12, 13]. Another well-known approach is the Fisherfaces in which the Fisher's Linear Discriminant (FLD) (i.e. LDA) is employed after PCA is used for dimensionality reduction [18]. Compared with the Eigenface (PCA) approach, the fisherface approach is more insensitive to large variations in lighting direction and facial expression. More recently, some variants of LDA (LDA) have been developed for face recognition such as F-LDA [23], D-LDA [21], FD-LDA [24] and KDDA [39] etc. However, the computational requirements of these approaches are greatly related to the dimensionality of the original data and the number of training samples. When the face database becomes larger, the time for training and the memory requirement will significantly increase. Moreover, the system based on PCA should be retrained when new classes are added. As a consequence, it is impractical to apply the PCA in systems with a large database.

Discrete Cosine Transform (DCT) has been employed in face recognition [44, 45, 85, 86]. The DCT has several advantages over PCA: (1) The DCT is data independent; (2) The DCT can be implemented using a fast algorithm. In [85], the DCT is applied for dimensionality reduction and then the selected low-frequency DCT coefficient vectors are fed into a multi-layer perceptron (MLP) classifier. It is well-known that the problems arising from the *curse of dimensionality* should be considered in pattern recognition. It has been suggested that as the dimensionality increases, the sample size needs to increase exponentially in order to have an effective estimate of multivariate densities [9]. In face recognition applications, the original input data are usually of high dimension, whereas only limited training samples are available. Therefore, dimensionality reduction is a very important

step which will greatly improve the performance of the face recognition system. However, if only the DCT is employed for dimensionality reduction, we cannot keep enough frequency components for important facial features in order to compromise the problem of *curse of dimensionality*. Besides, some variable features exist in the low dimensional features extracted by the DCT. Hence, in our proposed system, LDA is also employed in the DCT domain to extract the most discriminating features of face images. With the combination of DCT and LDA, more DCT coefficients can be kept and the most discriminating features can be extracted at high speed.

In [65], the clustering process is implemented after LDA is employed. However, LDA is a linear projection method and the results are globally optimal only for linearly separable data. There are lots of nonlinear variations in human faces such as pose and expression. Once the faces with different poses are put into the same class, they will actually smear the optimal projection such that we cannot get the most discriminating features and good generalization. For this reason, we should ensure that the face images in each class should not have large nonlinear variations. In this chapter, instead of regarding each individual as one class, we propose a sub-clustering method to split one class into several subclasses before implementing the LDA. Moreover, the sub-clustering process is very crucial to structure determination of the following RBF neural networks. In fact, the number of clusters is just the number of hidden neurons in the RBF neural networks.

Neural networks have been widely applied in pattern recognition for the reason that neural-networks-based classifiers can incorporate both statistical and structural information and achieve better performance than the simple minimum distance classifiers [1]. Multilayered networks (MLNs), usually employing the backpropagation (BP) algorithm, are widely used in face recognition [63]. Recently, RBF neural networks have been applied in many engineering and scientific applications including face recognition [65, 92]. The RBF neural networks possess the following salient features: 1) They are universal approximators [93]; 2) They have a simple

topological structure [94]; 3) They can implement fast learning algorithms because of locally tuned neurons [95]. Based on the advantages of RBF neural networks and the efficient feature extraction method, a high-speed RBF neural network classifier is proposed in this chapter. The near-optimal parameters can be estimated according to the properties of the feature space instead of using the gradient descent training algorithm. As a consequence, our system is able to achieve high training and recognition speed which facilitates real-time applications of the proposed face recognition system. The block diagram of our proposed face recognition system is shown in Fig. 4.1.

Figure 4.1: Block diagram of high-speed face recognition based on DCT and RBF neural networks

The organization of the chapter is as follows: In Section 4.2, we first present the DCT-based feature extraction method. Then, the proposed clustering algorithm and the LDA method implemented in the DCT domain are presented. Section 4.3 describes the architecture of RBF neural networks and the parameter estimation approach, which is based on the properties of previously extracted feature vectors. Experimental results are presented and discussed in Section 4.4. Finally, conclusions are drawn in Section 4.5.

4.2 Feature Extraction

4.2.1 Discrete Cosine Transform (DCT)

The Discrete Cosine Transform (DCT) has been widely applied to solve numerous problems among the digital signal processing community such as JPEG image

compression standard [87, 89]. The 2D $M \times N$ DCT is defined as follows:

$$C(u, v) = \alpha(u)\alpha(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cos\left[\frac{\pi(2x+1)u}{2M}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right] \quad (4.1)$$

and the inverse transform is defined as

$$f(x, y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \alpha(u)\alpha(v) C(u, v) \cos\left[\frac{\pi(2x+1)u}{2M}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right] \quad (4.2)$$

where

$$\alpha(u) = \begin{cases} \frac{1}{\sqrt{M}}, & u = 0 \\ \sqrt{\frac{2}{M}}, & u = 1, 2, \dots, M-1 \end{cases}$$

$$\alpha(v) = \begin{cases} \frac{1}{\sqrt{N}}, & v = 0 \\ \sqrt{\frac{2}{N}}, & v = 1, 2, \dots, N-1 \end{cases}$$

Figure 4.2: Scheme of scanning 2D DCT coefficients to a one-dimensional vector

In the JPEG image compression standard, original images are initially partitioned into rectangular non-overlapping blocks (8×8 blocks) and then the DCT is performed independently on the subimage blocks [89]. In our proposed system, we simply apply the DCT on the entire face image. If the DCT is only applied to the subimage independently, some relationship information between subimages cannot

be obtained. However, we can obtain all frequency components of a face image by applying the DCT on the entire face image. In addition, some low-frequency components are only related to the illumination variations which can be discarded. For an $M \times N$ image, we have an $M \times N$ DCT coefficient matrix covering all the spatial frequency components of the image. The DCT coefficients with large magnitude are mainly located in the upper-left corner of the DCT matrix. Accordingly, as illustrated in Fig. 4.2, we scan the DCT coefficient matrix in a zig-zag manner starting from the upper-left corner and subsequently convert it to a one-dimensional vector. Detailed discussions about image reconstruction errors using only a few significant DCT coefficients can be found in [85]. As a holistic feature extraction method, the DCT converts high dimensional face images into low dimensional spaces in which more significant facial features such as outline of hair and face, position of eyes, nose and mouth are maintained. These facial features are more stable than the variable high-frequency facial features. As a matter of fact, human visual system is more sensitive to variations in the low-frequency band. The limitation of the DCT for feature extraction is that it cannot efficiently utilize the higher-frequency features. Therefore, it is not capable of recognizing persons with similar looks. Nevertheless, for face recognition systems in which precise recognition is needed, it still can be employed as a course recognition step to make the high-speed recognition possible.

4.2.2 Clustering

Several well-known clustering algorithms such as k -means clustering and fuzzy k -means clustering are widely used in RBF neural networks [95, 96]. However, these clustering approaches are unsupervised learning algorithms and no category information about patterns is used. On the contrary, in face recognition, category information of training samples is previously known. Accordingly, we can simply cluster the training samples in each class. As depicted in Fig. 4.3, it is impossible for a purely unsupervised clustering algorithm to separate each class accurately.

Figure 4.3: Clustering with category information

In the proposed system, the classes are split in terms of their Euclidean distances. It is unreasonable to split the classes according to the degree of overlap because there are still great overlaps between classes after performing the DCT. Moreover, the following LDA will reduce the within-class scatter and thus make the sparsely distributed training samples in each subclass tighter. Hence, we only need to prevent the samples with large variations from being clustered in the same subclass. The proposed clustering algorithm is as follows:

- 1) Let u be the number of total clusters and s be the number of classes. Initially, set each class to be one cluster, i.e. $u = s$.
- 2) Find two training samples $x^k(f)$, $x^k(g)$ with the largest Euclidean distance $d^k(f, g)$ in class k , $k = 1, 2, \dots, s$. These two samples are called clustering reference sample (CRS).
- 3) Compute the Euclidean distances from the two samples to the samples $x^j(i)$ in other classes j , denoted as $d^{kj}(f, i)$ and $d^{kj}(g, i)$ respectively as follows:

$$d^{kj}(f, i) = \|x^j(i) - x^k(f)\|, \quad j \neq k \quad (4.3)$$

$$d^{kj}(g, i) = \|x^j(i) - x^k(g)\|, \quad j \neq k \quad (4.4)$$

where $\| \cdot \|$ denotes the Euclidean norm, $j = 1, 2, \dots, s$, $j \neq k$, $i = 1, 2, \dots, n^{\bar{k}}$, and $n^{\bar{k}}$ is the number of samples not belonging to class k .

- 4) Compute the mean value and the standard deviation of $d^{kj}(f, i)$, $d^{kj}(g, i)$ as follows:

$$\bar{d}^k(f) = \frac{1}{n^{\bar{k}}} \sum_{i=1}^{n^{\bar{k}}} d^{kj}(f, i) \quad (4.5)$$

$$v^k(f) = \left[\frac{1}{n^{\bar{k}}} \sum_{i=1}^{n^{\bar{k}}} (d^{kj}(f, i) - \bar{d}^k(f))^2 \right]^{\frac{1}{2}} \quad (4.6)$$

$$\bar{d}^k(g) = \frac{1}{n^{\bar{k}}} \sum_{i=1}^{n^{\bar{k}}} d^{kj}(g, i) \quad (4.7)$$

$$v^k(g) = \left[\frac{1}{n^{\bar{k}}} \sum_{i=1}^{n^{\bar{k}}} (d^{kj}(g, i) - \bar{d}^k(g))^2 \right]^{\frac{1}{2}} \quad (4.8)$$

- 5) Define the scope radius of the two CRSes $r^k(f)$, $r^k(g)$ as follows:

$$r^k(f) = \bar{d}^k(f) - \alpha v^k(f) \quad (4.9)$$

$$r^k(g) = \bar{d}^k(g) - \alpha v^k(g) \quad (4.10)$$

where α is a positive constant clustering factor. If $d^k(f, g) > \max\{r^k(g), r^k(f)\}$, then split the cluster into two clusters with CRSes $x^k(f)$ and $x^k(g)$ respectively and set $u = u + 1$.

- 6) There are three scenarios for any sample $x^k(h)$ in class k ($x^k(h)$ is not CRS). To make close data cluster together, these scenarios as depicted in Fig. 4.4 will be handled as follows:

- i) If only one CRS's scope comprises $x^k(h)$, then $x^k(h)$ will be merged with the cluster to which this CRS belongs.

- ii) If more than one CRS's scope comprise $x^k(h)$, then $x^k(h)$ will be merged into the cluster to which the CRS with the shortest distance to $x^k(h)$ belongs.
 - iii) If no CRS's scope comprise $x^k(h)$, then $x^k(h)$ is regarded as another CRS belonging to a new cluster. Set $u = u + 1$ and compute the radius $r^k(h)$ according to steps (4)-(5). Repeat step (6), until u dose not change.
- 7) Apply steps (2)-(6) to all classes.

Figure 4.4: Illustration of one class split into three subclasses

It follows from (4.9) and (4.10) that the radius of the CRS is chosen according to the mean distance and standard deviation from this CRS to the training samples belonging to other classes. The clustering factor α controls the clustering extent. The larger the value of α , the more clusters there are. Therefore, α should be chosen carefully so that the LDA will work efficiently. Fig. 4.5 shows the result of the proposed clustering algorithm. As depicted in Fig. 4.5, the two clusters are different in scale and pose. If these two clusters are put together, the LDA will not be able to efficiently obtain the discriminating features for this class. After the clustering algorithm and the LDA are implemented, the sparsely distributed training samples cluster more tightly which simplifies parameter estimation of the RBF neural networks in the sequel.

Figure 4.5: Example of training samples in one class splitted into two subclasses by the clustering algorithm ($\alpha = 1$)

4.2.3 Linear Discriminant Analysis (LDA)

In order to obtain the most salient and invariant features of human faces, LDA is applied in the truncated DCT domain. LDA is one of the most popular linear projection methods for feature extraction. It is used to find a linear projection of the original vectors from a high dimensional space to an optimal low dimensional subspace in which the ratio of the between-class scatter and the within-class scatter is maximized. In this chapter, we apply LDA to discount variations such as illumination and expression. The details about LDA can be found in Chapter 2. It should be noted that some important facial features will be discounted by LDA if face images with greatly different variations such as poses and expressions. As a consequence, we apply LDA after clustering such that the most discriminating facial feature can be effectively extracted. The discriminating feature vectors P projected from the truncated DCT domain to the optimal subspace can be calculated as follows:

$$P = E_{optimal}^T \cdot X \quad (4.11)$$

where X are truncated DCT coefficient vectors, and $E_{optimal}$ is the LDA optimal projection matrix.

4.3 Classification Using RBF Neural Networks

4.3.1 RBF Neural Networks

Figure 4.6: Architecture of RBF neural networks

The architecture of RBF neural networks is shown in Fig. 4.6. There are three layers in the traditional RBF neural networks, namely, input layer, hidden layer and output layer. Each layer is fully connected to the following layer. The hidden layer consists of RBF nodes with radial activation functions, which apply a nonlinear transformation from the input space to the hidden space. In the proposed system, we take the most frequently used Gaussian function as the radial basis function since it best approximates the distribution of data in each subset. In RBF neural networks, the RBF units are used to determine to which subclass an input vector belongs in terms of its distance from the center of the subclass.

Let $P \in \mathbb{R}^r$ be the input vector. The RBF unit is defined as follows:

$$R_i(P) = \exp\left(-\frac{\|P - C_i\|^2}{\sigma_i^2}\right) \quad i = 1, 2, \dots, u \quad (4.12)$$

where C_i is the center of the i th RBF unit, σ_i is the width of the i th RBF unit and u is the number of total RBF units.

The output layer implements a weighted sum of the RBF units as follows:

$$y_j(P) = \sum_{i=1}^u w(j, i) R_i(P) \quad (4.13)$$

The output layer is a linear combination of the radial basis functions. It forms linear decision boundaries in the output space. In fact, the RBF neural networks map the nonlinearly superable space into a space in which the data can be linearly separated [8].

4.3.2 Structure Determination and Parameter Estimation of RBF Neural Networks

In face recognition applications, the RBF neural networks are regarded as a mapping from the feature hyperspace to the classes. Therefore, the number of inputs of such an RBF neural network is determined by the dimension of input vectors. In the proposed system, the truncated DCT vectors after implementing LDA are fed to the input layer of the RBF neural network. The number of outputs is equal to the class number. The hidden neurons are very crucial to the RBF neural networks, which represent the subset of the input data. After the clustering algorithm is implemented, LDA projects the training samples into the subspace in which the training samples are clustered more tightly. Our experimental results show that the training samples are well separated and there are no overlaps between subclasses after LDA is performed. Consequently, in our system, the number of subclasses (i.e. the number of hidden neurons of the RBF neural networks) is determined by the previous clustering process.

In the proposed system, we simplify the estimation of the RBF parameters according to the data properties instead of supervised learning since the non-linear supervised method often suffers from having a long training time and the possibility of being trapped in local minima. Two important parameters are associated with each RBF unit, the center C_i and the width σ_i . Each center should well represent each subclass because the classification is actually based on the distances between the input samples and the centers of each subclass. There are different strategies in selecting RBF centers with respect to different applications [8]. Here, as LDA keeps the most discriminating feature for each sample in each subclass, it is reasonable to choose the mean value of the training samples in every subclass as the RBF center as follows:

$$C_i = \frac{1}{n^i} \sum_{j=1}^{n^i} P_j^i \quad (4.14)$$

where P_j^i is the j th sample in the i th subclass and n^i is the number of training samples in the i th subclass.

Width Estimation

To our knowledge, every subclass has its own features such that the individual scope of each subclass is different. The width of an RBF unit describes the properties of a subclass because the width of a Gaussian function represents the standard deviation of the function. Besides, the width controls the amount of overlap of Gaussian functions. If the widths are too large, there will be great overlaps between classes and thus the RBF units cannot represent the subclasses well and the output belonging to the class will not be so significant which will lead to great misclassifications. On the contrary, too small a width will result in rapid reduction in the value of a Gaussian function and thus poor generalization [97]. Accordingly, our goal is to select the width that minimizes the overlaps between different classes so as to preserve local properties while maximizing the generalization ability of the network.

As foreshadowed earlier, LDA makes the subclasses well separate. However, it has been indicated that the LDA method achieves the best performance on the training data, but generalizes poorly to new individuals, particularly when the training data set is small [98]. The distribution of training samples cannot well represent the new inputs. Hence, in this special case, the width of each subclass cannot be estimated merely according to the small number of training samples in each subclass. Our experimental studies show that the distances from the centers of RBF units to the new input samples belonging to other classes are similar to the distances to the training samples in other classes. These distances can be used to estimate the widths of RBF units since they generally imply the range of RBF units. In [99], it has been indicated that the patterns which are not consistent with data statistics (noisy patterns) should be rejected rather than used for training. Accordingly, the following method for width estimation is proposed:

$$dcc(j, i) = \| C_j^l - C_i^k \| \quad (4.15)$$

$$\begin{aligned} d_{med}(i) &= \text{med}\{dcc(j, i)\} \\ i, j &= 1, 2, \dots, u \quad i \neq j \quad k, l = 1, 2, \dots, s, \quad k \neq l \end{aligned} \quad (4.16)$$

where C_i^k is the center of the i th cluster belonging to the k th class and C_j^l is the center of the j th cluster belonging to the l th class and $d_{med}(i)$ is the median distance from the i th center to the centers belonging to other classes. In the proposed system, since the centers of RBF units well represent the training samples in each cluster, we estimate the width of one cluster by calculating the distances from this center to the centers belonging to other classes instead of the individual training samples so as to avoid excessive computational complexity. Hence, the

width σ_i of the i th cluster is estimated as follows:

$$\sigma_i = \frac{d_{med}(i)}{\sqrt{|\ln \gamma|}} \quad (4.17)$$

where γ is a factor that controls the overlap of this cluster with other clusters belonging to different classes. Equation (4.17) is derived from the Gaussian function. It should be noted that $d_{med}(i)$ is determined by the distances to the cluster centers belonging to other classes (not other clusters) because one class can be split into several clusters and the overlaps between clusters from the same class are allowed to be large. The median distance $d_{med}(i)$ well measures the relative scope of RBF units. Furthermore, by selecting the proper factor γ , suitable overlaps between different classes can be guaranteed.

Weight Adjustment

In the first stage, we estimate the parameters of the RBF units by using unsupervised training methods. The second phase of training is to optimize the second-layer weights of the RBF neural networks. Since the output of the RBF neural networks is a linear model, we can apply linear supervised learning to minimize a suitable error function. The sum-of-squares error function is given by

$$E = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^s (t_i^j - y^j(P_i))^2 \quad (4.18)$$

where t_i^j is the target value for output unit j when the i th training sample P_i is fed to the network, $y^j(P_i) = \sum_{k=1}^u w(j, k)R_k$, R_k is the k th output of the RBF unit, u is the number of RBF units generated according to the clustering algorithm in Section 4.2.2 and n is the total number of training samples. This problem can be solved by the linear least square (LLS) paradigm [9].

Let r and s be the number of input and output neurons respectively. Furthermore, let $R \in \mathbb{R}^{u \times n}$ be the RBF unit matrix and $T = (T_1, T_2, \dots, T_n)^T \in \mathbb{R}^{s \times n}$ be the target matrix consisting of “1’s” and “0’s” with exactly one per column that

identifies the processing unit to which a given exemplar belongs. The problem of finding an optimal weight matrix $W^* \in \mathbb{R}^{s \times u}$ such that the error function (4.18) is minimized as follows:

$$W^* = (TR^\dagger)^T \quad (4.19)$$

where $R^\dagger$ is the pseudoinverse of R and is given by

$$R^\dagger = (R^T R)^{-1} R^T \quad (4.20)$$

In the proposed system, however, direct solution of (4.20) can lead to numerical difficulties due to the possibility of $R^T R$ being singular or near singular. This problem can be best solved by using the technique of singular value decomposition (SVD) [100]. Suppose the singular value decomposition of R is $R = U \Sigma V^T$. The pseudoinverse of R is

$$R^\dagger = V \Sigma^\dagger U^T \quad (4.21)$$

where Σ is a $u \times n$ matrix with singular values $\sigma_1, \sigma_2, \dots, \sigma_q$ on its diagonal, and the reciprocals $1/\sigma_1, 1/\sigma_2, \dots, 1/\sigma_q$ are on the diagonal of $\Sigma^\dagger$ ($n \times u$).

4.4 Experimental Results and Discussion

In order to evaluate the proposed face recognition system, our experiments are performed on two benchmark face databases: 1) The ORL database; 2) The Yale database. Besides, for each database, we use three different evaluation methods which are mostly used in each database respectively. In this way, the experiment results can be compared with other face recognition approaches fairly.

4.4.1 Testing on the ORL Database

In this section, our face recognition system is tested on the ORL database. The ORL database contains a set of face images taken between April 1992 and April

1994 at the Olivetti Research Laboratory (ORL) in Cambridge University, UK. There are 400 images of 40 individuals. For some subjects, the images were taken at different times, which contain quite a high degree of variability in lighting, facial expression (open/closed eyes, smiling/non-smiling etc), pose (up-right, frontal position etc), and facial details (glasses/no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position, with tolerance for some tilting and rotation of up to 20°. The variation in scale is up to about 10%. A preview image of the database is shown in Fig. 4.7.

Figure 4.7: The ORL face database

In the following experiments, 5 images are randomly selected as the training samples and another 5 images as test images. Therefore, a total of 200 images are used for training and another 200 for testing and there are no overlaps between the training and testing sets. Here, we verify our system based on the average error rate, E_{ave} , which is defined as

$$E_{ave} = \frac{\sum_{i=1}^q n_{mis}^i}{qn_t} \quad (4.22)$$

where q is the number of simulation runs. (The proposed system is evaluated on ten runs, $q = 10$.), n_{mis}^i is the number of misclassifications for the i th run and n_t is the total number of testing samples for each run. We also denote the maximum and minimum misclassification rates for the ten runs as E_{max} and E_{min} respectively.

The dimension of feature vectors fed into RBF neural networks is essential for recognition. In [65], experimental results showed that the best results are achieved when the feature dimension is 25-30. If the feature dimension is too small, the feature vectors do not contain sufficient information for recognition. However, it is not guaranteed that more information will result in higher recognition rate. It has been indicated that if the dimension of the network input is comparable to the size of the training set, the system is liable to overfitting and result in poor generalization [64]. Moreover, some unimportant information may become noise and degrade the performance. Our experiments also showed that the best recognition rates are achieved when the feature dimension is about 30. Hence, the feature dimension of 30 will be adopted in the following simulation studies.

Parameter Selection

Two parameters, namely the clustering factor α and the overlapping factor γ need to be determined. As foreshadowed in Section 4.2.2, the sub-clustering process is based on the mean value and standard deviation of the distances from the CRS to the samples in other classes. Normally, we can choose $\alpha = 1$ as the initial value since the difference between the mean distance and standard deviation

approximately implies the scope of CRS. Nevertheless, with different databases and applications, the proper value of α should be adjusted accordingly. For the ORL database, the following experimental results show that the proper value of α lies in the range of $1 \leq \alpha \leq 2$.

Since the overlapping factor γ is not related to α , we can fix the value of α when estimating γ . The value of α is set to 1 for the following parameter estimation process. It follows from (4.17) that the factor γ actually represents the output of the RBF unit when the distance between the input and the RBF center is equal to d_{med} . As a result, γ should be a small value. We can initially assume that γ lies in the range of $0 < \gamma < 0.3$. More optimal and precise γ can be further estimated by finding the minimum value of the root mean square error (RMSE). The RMSE curves for five different training sets are depicted in Fig. 4.8. It is evident that there is only one minimum value in each RMSE curve and it usually lies in the range of $0.05 \leq \gamma \leq 0.1$. We should not choose the exact minimum value of γ because LDA makes the training samples in each cluster tighter in comparison with the testing samples. Accordingly, in order to obtain better generalization, we choose a slightly larger value for γ . In the following experiments, we choose the value of 0.1 for the overlapping factor γ which is shown to be a proper value for the RBF width estimation in our system. In some applications, parameter selection can be based on the cross-validation method to achieve good generalization ability.

Number of DCT Coefficients

In order to determine how many DCT coefficients should be chosen, we test the recognition performance with different number of DCT coefficients. Simulation results are summarized in Table 4.1. Here, the clustering coefficient α is set to 1. We can see from Table 4.1 that more DCT coefficients do not necessarily mean better recognition performance because high-frequency components are related to unstable facial features such as expression. There will be more variable information

Figure 4.8: RMSE curve

for recognition when the DCT coefficients increase. According to Table 4.1, the best performance is obtained when 50-60 DCT coefficients are used in our recognition system. In addition, Table 4.1 shows that the performance of our system is relatively stable when the number of DCT coefficients changes significantly. This is mainly due to the LDA algorithm which discounts the irrelevant information while keeping the most invariant and discriminating information for recognition.

Table 4.1: Recognition performance versus number of DCT coefficients ($\alpha = 1$ $\gamma = 0.1$)

No. of DCT coefficients	Feature dimension	$E_{min}(\%)$	$E_{max}(\%)$	$E_{ave}(\%)$
35	30	1.5	10.0	4.55
40	30	1.5	7.0	3.90
45	30	1.5	8.5	3.85
50	30	2.0	7.5	3.60
55	30	1.0	6.5	3.20
60	30	1.5	8.0	3.30
70	30	2.5	9.0	4.05
80	30	1.5	8.0	3.80
90	30	2.0	9.0	4.90
100	30	3.5	10.0	5.50

Effect of Clustering

As mentioned in Section 4.2, LDA is a linear projection paradigm and it cannot handle nonlinear variations in each class. Therefore, the proposed sub-clustering algorithm is applied before taking LDA. The clustering factor α controls the extent of clustering and determines the number of RBF units. Small number of clusters may lead to great overlap between classes and cannot obtain the optimal LDA projection direction. On the other hand, an increase of clusters may result in poor generalization because of overfitting. Moreover, since the training samples in each class are limited, the increase of clusters leads to reduction of training samples in each cluster so that LDA will work inefficiently. Table 4.2 shows one run of the recognition results with different numbers of clusters where E denotes the misclassification rate. The best performance is obtained when α lies in the range of 1.0-2.0. The results show that sub-clustering will improve the performance even when LDA is applied on small clusters. Without the clustering process, face images with large nonlinear variations will be in the same cluster. LDA will discount some important facial features instead of extracting them since LDA is a linear global projection method. Therefore, for face images with large variations such as pose, scale etc., the sub-clustering is necessary before implementing LDA. This process will be more effective if there are more training samples for each cluster.

Table 4.2: Recognition performance versus clustering factor α ($\gamma = 0.1$)

No. of DCT coefficients	Feature dimension	No. of clusters	α	$E(\%)$
55	30	40	0.0	4.0
55	30	42	0.5	3.0
55	30	44	1.0	3.0
55	30	46	1.2	3.0
55	30	50	1.5	2.5
55	30	55	1.8	3.0
55	30	58	2.0	3.0
55	30	74	2.5	4.0
55	30	107	3.0	4.5

By setting the optimal parameters in the proposed system, we obtain high recogni-

tion performance based on ten simulation studies whose results are shown in Table 4.3 (σ_E is the standard deviation for the mean of the error rate).

Table 4.3: Performance on 10 simulations ($\alpha = 1.5, \gamma = 0.1$)

No. of DCT coefficients	Feature dimension	$E_{min}(\%)$	$E_{max}(\%)$	$E_{ave}(\%)$	$\sigma_E(\%)$
55	30	0.0	4.5	2.45	0.45

Comparisons with Other Approaches

Many face recognition approaches have been performed on the ORL database. In order to compare the recognition performance, we choose some recent approaches tested under similar conditions for comparison. Approaches are evaluated on recognition rate, training time and recognition time. Comparative results of different approaches are shown in Table 4.4. Our experiments are performed on a Pentium II 350M computer, using Windows 2000 and Matlab 6.1. It is hard to compare the speed of different algorithms performed on different computing platforms. Nevertheless, we can approximately compare their relative speeds based on the normalized training and recognition time listed in Table 4.4. The time is normalized based on 1000M CPU frequency. It is evident from the Table that our proposed approach achieves high recognition rate as well as high training and recognition speed.

Computational Complexity

In this section, in order to provide more information about the computational efficiency of the proposed system, the approximate complexity of each part is analyzed and the results are summarized in Table 4.5.

In face recognition applications, the dimensionality of an original face image is usually considerably greater than the number of training samples. Therefore, the computational complexity mostly lies in the dimensionality reduction stage. The

Table 4.4: Recognition performance comparison of different approaches

Approach	Error rate(%)		Normalized training t	Normalized recognition t	Platform
	Best	Mean			
Pseudo-2D HMM [42]	5.0	n/a	n/a	16.8 sec.	Sun SparcII
SOM+CN [59]	1.5	3.8	24 min.	0.05 sec.	SGI Indy100M
PDBNN [61]	n/a	4.0	2 min.	0.01 sec.	SGI Indy 100M
DCT+HMM [45]	0.5	n/a	4.7 sec.	0.7 sec.	Pentium200M
Pseudo-2D HMM+DCT [44]	0.0	n/a	n/a	0.6 sec.	Pentium400M
LVQ+RBF+FEC [101]	0.5	n/a	n/a	0.31 sec.	Intel200M
DCT+MLP [85]	3.0	5.8	27 sec.	0.00009 sec.*	IBM450M
FND [102]	n/a	1.75	24.4 sec.	1.29 sec.	PentiumII400M
NFA-II+NFL [103]	n/a	4.05	n/a	0.272 sec.	IBM1700M
Our proposed approach	0.0	2.45	4.9 sec.	0.019 sec.	PentiumII350M

* It is not clear if the computational time for the DCT is counted in because the DCT takes about 0.016 seconds per image in our system. The time for classification is only about 0.00031 seconds.

training and recognition speed are greatly improved because the fast DCT reduces the computational complexity from $O(N^4)$ to $O(N^2 \log N)$ for an $N \times N$ image where N is a power of 2 [87]. Moreover, the proposed parameter estimation method is much faster than the gradient descent training algorithm which will take up to hundreds of epochs.

Performances with Different Number of Training Samples

Since LDA is a statistical method for feature extraction, the choice of training samples will affect its performance. In [19], the authors indicate that LDA works efficiently only when the number of training samples is large and representative for each class. Moreover, small number of training samples will result in poor generalization for each RBF unit. It will be difficult for a face system to recognize a face with too large variations such as pose and expression. Therefore, any possible training sample should be used such that better generalization can be achieved. That is also the reason why online training is essential for a face recognition system with high performance. Simulation results with different number of training samples are shown in Fig. 4.9. Our approach is promising if more training samples are available.

Table 4.5: Computational complexity

Part	Complexity
DCT	$O(N^2 \log N N_{tr})$
Clustering	$O(N_{dct} N_{tr}^2)$
LDA	$O(N_u N_{dct} + N_r^3 + N_{dct} N_r N_{tr})$
RBF parameter estimation	$O(N_u^3)$
LLS	$O(N_s N_u N_n + N_n^2 N_s + N_{tr}^2 \log N_{tr} + N_n N_s^2)$
Classification	$O(N^2 \log N + N_{dct} N_r + N_u N_s)$

- N The dimension of an $N \times N$ face image (N is a power of 2) ^a
- N_{tr} The number of training samples
- N_{dct} The number of truncated DCT coefficients
- N_r The number of input neurons (The dimension of the LDA feature vectors)
- N_u The number of clusters (The number of RBF units)
- N_s The number of output neurons (The number of classes)

^aThe original image can be padded with zeros to the size of $N \times N$

4.4.2 Testing on the Yale Database

We further test our proposed approach on the Yale database which contains 165 frontal face images covering 15 individuals taken under 11 different conditions. Each individual has different facial expressions (happy, sad, winking, sleepy, and surprised), illumination conditions (center-light, left-light, right-light) and small occlusion (with glasses). In order to be consistent with the experiments of [18], the centered face images of the normalized Yale face database and their closely cropped images are used in our experiment.[†] Examples of the centered Yale images and the closely cropped Yale images are shown in Fig. 4.10 and Fig. 4.11 respectively.

For the Yale face database, our system is also tested by using the “leaving-one-out” strategy which was used in [18]. In order to fairly evaluate different feature extraction methods, the nearest neighbor classifier was used for classification. We also skip the sub-clustering process since only frontal normalized face images are used. Comparative results are shown in Table 4.6. The results of Eigenface (PCA)

[†]The normalized Yale face database can be obtained from <http://www-white.media.mit.edu/vismod/classes/mas622-00/datasets/>

Figure 4.9: Performances with different number of training samples (Results are based on ten runs)

Figure 4.10: Examples of the centered Yale images

and Fisherface in our experiment are not identical to the results shown in [18]. This is mainly due to little different cropping size and normalization. Nevertheless, the results are very close. We can see from Table 4.6 that our feature extraction method achieves similar performance compared to the Fisherface method. However, the main advantage of our method is its low computational load which is crucial for high-speed face recognition in large databases. If we use the proposed RBF neural networks for classification, our system achieves better performance. (See Table 4.7).

Figure 4.11: Examples of the closely cropped Yale images

Table 4.6: Comparison on the Yale database

Approach	Reduced space	Error rate (%)	
		Close crop	Full face
Eigenface	30	25.5	17.6
Eigenface w/o 1st 3	30	17.6	12.1
Fisherface	15	6.1	3.0
DCT	55	26.1	20.0
DCT + LDA	15	6.7	3.6

4.5 Conclusion

This chapter presents a high-speed face recognition method based on the techniques of DCT, LDA and RBF neural networks. Facial features are first extracted by the DCT which greatly reduces dimensionality of the original face image as well as maintains the main facial features. Compared with the well-known PCA approach, the DCT has the advantages of data independency and fast computational speed. In order to obtain the most invariant and discriminating feature of faces, the linear projection method LDA is further applied to the feature vectors. Before implementing LDA, we propose a clustering algorithm to prevent training samples with large variations from being clustered in the same class. This process guarantees optimal projection direction for LDA and determines the number of hidden neurons of the RBF neural networks for classification. After LDA is applied, there are no overlaps between classes and the architecture and parameters of RBF neural networks are determined according to the distribution properties of the

Table 4.7: Performance on the Yale database ($\gamma = 0.1$)

Approach	Error rate (%)	
	Close crop	Full face
DCT + LDA + RBF	4.8	1.8

training samples. Simulation results on two benchmark face databases show that our system achieves high training and recognition speed, as well as high recognition rate.

Chapter 5

Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithm Domain

This chapter presents a novel illumination normalization approach for face recognition under varying lighting conditions. In the proposed approach, Discrete Cosine Transform (DCT) is employed to compensate for illumination variations in the logarithm domain. Since illumination variations mainly lie in the low-frequency band, an appropriate number of DCT coefficients are truncated to minimize variations under different lighting conditions. Experimental results on the Yale B database and CMU PIE database show that the proposed approach improves the performance significantly for the face images with large illumination variations. Moreover, the advantage of our approach is that it does not require any modelling steps and can be easily implemented in a real-time face recognition system.

5.1 Introduction

Recently, more researchers focus on robust face recognition such as face recognition systems invariant to pose, expression and illumination variations. Illumination variation is still a challenging problem in face recognition research area, especially for appearance-based approaches. The same person can appear greatly different under varying lighting conditions. A variety of approaches have been proposed to solve the problem [18, 79, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113]. These approaches can be generally classified into three main categories:

Preprocessing and Normalization: In this approach, face images are preprocessed using some image processing techniques to normalize the images to appear stable under different lighting conditions. For instance, histogram equalization (HE), Gamma correction and logarithm transform etc. are widely used for illumination normalization [104, 105]. However, non-uniform illumination variation is still difficult to deal with using these global processing techniques. Recently, adaptive histogram equalization (AHE) [114], region-based histogram equalization (RHE) [104] and block-based histogram equalization (BHE) [106] have also been proposed to cope with non-uniform illumination variations. Although recognition rates on face databases with non-uniform illumination variations can be improved compared with the HE, their performance results are still not satisfactory. In [111], by combining symmetric shape-from-shading (SSFS) and a generic 3D model, the performance of face recognition under varying illuminations is enhanced. However, this method is only efficient for exact frontal face images and it is assumed that all faces share a similar common shape. In [104], the authors proposed a normalization method called Quotient Illumination Relighting (QIR). This method is based on the assumption that the lighting modes of the images are known or can be estimated.

Invariant Feature Extraction: This approach attempts to extract facial features which are invariant to illumination variations. Edge maps, derivatives of

the gray-level and Gabor-like filters are investigated in [108]. However, empirical studies show that none of these representations are sufficient to overcome image variations due to changes in the direction of illumination. Another well-known feature extraction method is called Fisherface (also known as Linear Discriminant Analysis (LDA)) which linearly projects the image space to a low dimensional subspace to discount variations in lighting and facial expressions [18]. But, this method is a statistical linear projection method which largely relies on representativeness of the training samples. In [110], the Quotient Image is regarded as the illumination invariant signature image which can be used for face recognition under varying lighting conditions. Bootstrap database is required for this method and the performance degrades when dominant features between the bootstrap set and the test set are misaligned.

Face Modelling: Illumination variations are mainly due to the 3D shape of human faces under lighting in different directions. Recently, some researchers attempt to construct a generative 3D face model that can be used to render face images with different poses and under varying lighting conditions [79, 107, 109, 112]. A generative model called illumination cone was presented in [79, 107]. The main idea of this method is that the set of face images in fixed pose but under different illumination conditions can be represented using an illumination convex cone. The illumination convex cone can be constructed from a number of images acquired under variable lighting conditions and it can be approximated in a low dimensional linear subspace. In [109], the authors showed that the set of images of a convex Lambertian object obtained under a variety of lighting conditions can be well approximated by a 9D linear subspace. One of the drawbacks of the model-based approaches is that a number of images of the subject under varying lighting conditions or 3D shape information are needed during the training phase. This drawback limits their applications in practical face recognition systems. In addition, existing model-based approaches assume that the human face is a convex object, i.e., the casting shadows are not considered. The specular problem is also ignored even though the human face is not a perfect Lambertian surface.

To the best of our knowledge, one ideal way of solving the illumination variation problem is to normalize a face image to a standard form under uniform lighting conditions. In fact, the human visual system usually cares about the main features of a face, such as the shapes and relative positions of the main facial features, and ignores illumination changes on the face while recognizing a person. Accordingly, in this chapter, we propose an illumination normalization approach to remove illumination variations while keeping the main facial features unimpaired. The key idea of the proposed approach is that illumination variations can be significantly reduced by truncating low-frequency DCT coefficients in the logarithm DCT domain. Our approach can be categorized into the first approach group although feature extraction can be carried out directly in the logarithm DCT domain.

The remainder of this chapter is organized as follows. In Section 5.2, we describe the illumination normalization approach in the logarithm DCT domain in detail. Experimental results and discussions are presented in Section 5.3. Finally, conclusions are drawn in Section 5.4.

5.2 Illumination Normalization in the Logarithm DCT Domain

5.2.1 Logarithm Transform

Logarithm transform is often used in image enhancement to expand the values of dark pixels [91, 108]. Here, we show why illumination compensation should be implemented in the logarithm domain. In a simple situation, the image gray level $f(x, y)$ can be assumed to be proportional to the product of the reflectance $r(x, y)$ and the illumination $e(x, y)$ [115], i.e.

$$f(x, y) = r(x, y) \cdot e(x, y) \quad (5.1)$$

To our knowledge, the Retinex algorithm is related to the reflectance constancy [116]. The invariant property of reflectance ratio has been applied in object recognition [117]. Since the reflectance is a stable characteristic of facial features, our goal is to recover the reflectance of faces under varying illumination conditions. Taking logarithm transform on (5.1), we have:

$$\log f(x, y) = \log r(x, y) + \log e(x, y) \quad (5.2)$$

It follows from (5.2) that in the logarithm domain, if the incident illumination $e(x, y)$ and the desired uniform illumination e' are given (e' is identical for every pixel of an image), we have:

$$\begin{aligned} \log f'(x, y) &= \log r(x, y) + \log e' \\ &= \log r(x, y) + \log e(x, y) - \epsilon(x, y) \\ &= \log f(x, y) - \epsilon(x, y) \end{aligned} \quad (5.3)$$

where $\epsilon(x, y) = \log e(x, y) - \log e'$ and $f'(x, y)$ is the pixel value under desired uniform illumination. From (5.3), we can conclude that the normalized face image can be obtained from the original image by using an additive term $\epsilon(x, y)$ called compensation term which is the difference between the normalized illumination and the estimated original illumination in the logarithm domain.

5.2.2 Illumination Compensation

Given a face image, illumination variations can be well compensated by adding or subtracting the compensation term $\epsilon(x, y)$ of (5.3) in the logarithm domain if we know where illumination variations and important facial features are. However, facial feature detection is a non-trivial task especially for face images with large illumination variations. According to the Lambertian reflectance theory, illumination variations are mainly caused by the shape of a face. Normally, the surface

of a face does not change abruptly. Illumination changes slowly compared with the reflectance of facial features except some casting shadows and specularities on the face. As a result, illumination variations mainly lie in the low-frequency band. Since we attempt to recognize faces using reflectance characteristic, illumination variations can be reduced by removing low-frequency components. It should be noted that only face images without hair are considered in our approach because the intensity of human's hair is a kind of low-frequency feature which will be impaired by discarding low-frequency components of face images. However, human's hair is a kind of unstable feature which will change greatly with time. Therefore, in many face recognition systems, human's hair is not regarded as a kind of important facial feature.

In the proposed approach, the DCT is performed on the entire face image to obtain all frequency components of the face image. The DCT can be used to transform an image from spatial domain to frequency domain. Besides, it can be implemented using a fast algorithm which significantly reduces the computational complexity. Low-frequency components of a face image can be removed simply by setting the low-frequency DCT coefficients to zero. Evidently, the resulting system works like a high-pass filter. Since illumination variations are mainly low-frequency components, we can estimate the incident illumination on a face by using low-frequency DCT coefficients. It follows from (4.1) that setting the DCT coefficients to zero is equivalent to subtracting the product of the DCT basis image and the corresponding coefficient from the original image. If n low-frequency DCT coefficients are set to zero, we have

$$\begin{aligned}
 F'(x, y) &= \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} E(u, v) - \sum_{i=1}^n E(u_i, v_i) \\
 &= F(x, y) - \sum_{i=1}^n E(u_i, v_i)
 \end{aligned} \tag{5.4}$$

where

$$E(u, v) = \alpha(u)\alpha(v)C(u, v) \cos\left[\frac{\pi(2x+1)u}{2M}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right]$$

Since illumination variations are expected to be in the low-frequency components, the term $\sum_{i=1}^n E(u_i, v_i)$ can be approximately regarded as the illumination compensation term. It follows from (5.3) that the term $F'(x, y)$ in (5.4) is just the desired normalized face image in the logarithm domain. Therefore, discarding low-frequency DCT coefficients in logarithm domain is equivalent to compensating for illumination variations. This is the reason why DCT should be implemented in the logarithm domain.

The first DCT coefficient (i.e., the DC component) determines the overall illumination of a face image. Therefore, the desired uniform illumination can be obtained by setting the DC coefficient to the same value, i.e.

$$C(0, 0) = \sqrt{MN} \cdot \log \mu \quad (5.5)$$

where $C(0, 0)$ is the DC coefficient of the logarithm image. For the convenience of understanding and visualization, we normally choose a value of μ near the middle level of the original image. In other words, the normal face has an average gray level of μ . It should be noted that we do not regard the skin color as a kind of facial feature because it is unstable when illumination changes. For example, the face of a black man normally has an average gray level below μ . It is actually regarded as a normal face under weak illumination conditions. It follows from (5.3) and (5.4) that the difference between the original DC component and the normalized DC component, together with the other discarded low-frequency AC components, approximately make up the compensation term $\epsilon(x, y)$.

5.2.3 Logarithm Domain versus Original Domain

Since illumination variations mainly lie in the low-frequency band, we can approximately estimate them using the low-frequency DCT basis images and their corresponding coefficients. Fig. 5.1 shows the first several DCT basis images.

Figure 5.1: DCT basis images.

Figure 5.2: (a) Original image; (b) Reconstructed image by applying DCT on the original image and discarding the (0,1)th DCT coefficient; (c) Reconstructed image by applying DCT on the logarithm image and discarding the (0,1)th DCT coefficient ($\mu = 100$); (d) Reconstructed logarithm image by applying DCT on the logarithm image and discarding the (0,1) DCT coefficient (i.e., (c) without the inverse logarithm transform).

As a simple example to illustrate the idea, note that the half-lighted face image is highly correlated with the (0,1)th basis image. In other words, the illumination difference on a half-lighted face can be approximately estimated from the (0,1)th DCT coefficient. Therefore, the invariant reflectance can be obtained by discarding

the $(0, 1)$ th DCT coefficient. As illustrated in Fig. 5.2, the facial features in the dark area of the original image are recovered much better by applying DCT on the logarithm image. In fact, only the brightness of the image is adjusted by discarding DCT coefficients of the original image, whereas discarding DCT coefficients of the logarithm image will adjust the illumination and recover the reflectance characteristic of the face.

5.2.4 Logarithm Image for Recognition

Human faces are not perfect Lambertian surfaces [109]. In some cases, there are specularities on a face image which do not lie in the low-frequency band. Moreover, some shadows also lie in the same frequency band as the main facial feature. As a consequence, illumination variations on some small areas may not be correctly compensated by discarding the low-frequency coefficients. For example, some small areas under high illumination level may be incorrectly adjusted to even higher level. As we know, the addition operation in the logarithm domain is equivalent to multiplication in the original domain. If the logarithm image is restored to the original one, incorrect adjustment will make it even worse. Accordingly, in our approach, logarithm images are directly used for recognition, i.e., the inverse logarithm transform step is skipped. In fact, there are also physiological evidences that the response of the retina cells can be approximated as a log function of the intensity [108]. Fig. 5.2(c),(d) show the reconstructed non-logarithm image and the logarithm image after discarding the $(0, 1)$ th DCT coefficient to correct half-lighted illumination.

5.2.5 Discarding DCT Coefficients

As aforementioned, low-frequency DCT coefficients which are highly related to illumination variations should be discarded. There remains another issue: which and how many DCT coefficients should be discarded in order to obtain the well

normalized face image?

Figure 5.3: Standard deviations of the logarithm DCT coefficients.

Figure 5.4: Manner of discarding DCT coefficients.

Fig. 5.3 shows the standard deviations of the logarithm DCT coefficients which are calculated from 64 face images of the same subject (only the first 30×30 coefficients are shown). As we can see from Fig. 5.3, standard deviations of the coefficients with great magnitude are mainly located in the upper-left corner of the DCT coefficient matrix. Accordingly, illumination variations of face images can be reduced by discarding these low-frequency coefficients (The DC coefficient is set to a constant value according to (5.5)). The manner of discarding DCT coefficients is shown in Fig. 5.4.

5.3 Experimental Results and Discussion

5.3.1 Face Database

In this chapter, the Yale Face Database B and the CMU PIE Face Database are both used to evaluate the proposed approach. These two face databases contain face images with large illumination variations.

Yale Face Database B

There are 10 individuals under 64 different lighting conditions for 9 poses in the database. Since we are only concerned with the illumination problem in this chapter, frontal face images under varying lighting conditions are used. As shown in Table 5.1, the face images are divided into 5 subsets according to the angle between the light source direction and the camera axis. Interested readers may refer to [79] for more detailed information of the database. In the experiments, face images are all cropped and aligned in accordance with [79] and all the face images are rescaled to the size of 120×105 . Fig. 5.5 shows the images of one individual divided into five subsets based on different lighting conditions. We use Subset 1 as the training set and other subsets are used for testing.

Table 5.1: Subsets divided according to light source direction

Subsets	1	2	3	4	5
Lighting angle ($^{\circ}$)	$0 \sim 12$	$13 \sim 25$	$26 \sim 50$	$51 \sim 77$	> 77
Number of images	70	120	120	140	190

CMU PIE Face Database

In the CMU PIE database, there are 68 subjects with pose, illumination and expression (PIE) variations. In our experiments, only frontal face images under different lighting conditions (without expression variations) are selected [80]. As shown in Fig. 5.6, the frontally lighted face images are chosen as the training set.

Figure 5.5: Sample images of an individual divided into 5 subsets.

The remaining 20 face images of each subject with different illumination variations are used for testing. Face images in the experiment are all cropped and aligned in the same way as images in the Yale B database.

5.3.2 Experimental Results

In the experiments, the nearest neighbor classifier based on the Euclidean distance is employed for classification. All the face images used in the experiments are normalized so that they have zero mean and unit variance.

Figure 5.6: Sample images of an individual in CMU PIE database: (a) Training; (b) Testing.

An appropriate number of discarded DCT coefficients should be chosen in order to normalize the illumination well and not weaken important facial features. We employ the dimensionality of the discarded coefficients (D_{dis} shown in Fig. 5.4) to measure the extent of discarding coefficients. Fig. 5.7 shows examples of normalized logarithm images with different D_{dis} .

Results of employing correlation and Eigenface methods [13] on the normalized face images of both databases with different D_{dis} are shown in Fig. 5.8, Fig. 5.9 and Fig. 5.10. For the Eigenface method, 50 principal components are used. It is evident that the error rate significantly decreases after a few DCT coefficients are discarded. As illustrated in Fig. 5.8, Fig. 5.9 and Fig. 5.10, the best and

Figure 5.7: Normalized logarithm images with different D_{dis} : (a)Original image; (b) $D_{dis} = 3$; (c) $D_{dis} = 6$; (d) $D_{dis} = 15$; (e) $D_{dis} = 20$; (f) $D_{dis} = 35$; (g) $D_{dis} = 50$.

stable performances approximately lie in the range of $18 \leq D_{dis} \leq 25$. In other words, in this range, illumination variations are largely reduced while important facial features are preserved.

For the Yale B database, the small number of subjects is one of the reasons that the performance does not drop even when D_{dis} is around 50 because the high-frequency features are enough to distinguish these few subjects. Another reason is that discarding low-frequency DCT coefficients keeps high-frequency features well. As a matter of fact, illumination variations and facial features are not perfectly separated with respect to frequency components. Some illumination variations, especially shadows and specularities, lie in the same frequency bands as some facial features do. As a consequence, in order to compensate for such illumination variations, some facial information, mainly low-frequency intensity variations of facial feature components, has to be sacrificed. Nevertheless, our experiments show that high performance can still be achieved without these features. The low-frequency features actually become less effective under large illumination variation conditions. From Fig. 5.8, Fig. 5.9 and Fig. 5.10, we can see that the performance has significantly improved when $D_{dis} > 5$. Therefore, for some applications without large illumination variations, especially shadowing, more low-frequency components can be preserved. It should be noted that our method is different from the DCT applied

for dimensionality reduction in face recognition [86], in which only low-frequency coefficients are used as facial features. Our method should use higher frequency features in order to reduce the illumination variations. Nevertheless, if logarithm images are used for recognition based on their method, robustness against illumination variations can be easily improved by discarding several low-frequency DCT coefficients.

Figure 5.8: Performance on the Yale B database with different D_{dis} (Correlation).

Figure 5.9: Performance on the Yale B database with different D_{dis} (Eigenfaces).

Comparison results with other methods dealing with illumination variations on both databases (mainly Yale B) are shown in Table 5.2. The results of our normalization method are the average error rates of $18 \leq D_{dis} \leq 25$. Some listed

Figure 5.10: Performance on the CMU PIE database with different D_{dis} .

results of the existing methods are directly from other papers since they are based on the same database. We can see from Table 5.2 that the proposed method outperforms most of the existing methods except the cones-cast method. However, it should be pointed out that the illumination cone method needs much more complicated modelling steps, thus it cannot be applied in some practical applications. Moreover, in their paper, results on the most difficult subset (Subset 5) are not given.

Table 5.2: Recognition performance comparisons of different methods

Methods	Error rate (%)			
	Yale B			CMU PIE
	Subset 3	Subset 4	Subset 5	
No normalization	10.8	51.4	77.4	43.0
Histogram equalization	9.2	54.2	41.1	47.8
Linear subspace [79]	0	15.0	n/a	n/a
Cones-attached [79]	0	8.6	n/a	n/a
Cones-cast [79]	0	0	n/a	n/a
Harmonic images [112]	0.3	3.1	n/a	n/a
Illumination ratio images [113]	3.3	18.6	n/a	n/a
Quotient illumination relighting [104]	0	9.4	17.5	n/a
Our method	0	0.18	1.71	0.36

5.3.3 Performance Comparison between Logarithm and Non-logarithm Images

As described in Section 5.2.4, for face recognition, normalized logarithm face images should outperform non-logarithm (i.e. with inverse logarithm transform) face images. Performance comparisons using the correlation method on the Yale B (Subset 4 and 5) and CMU PIE databases are shown in Fig. 5.11, Fig. 5.12 and Fig. 5.13 respectively. It is clear that better performance is achieved while using logarithm face images. It is more evident in the initial stage of discarding DCT coefficients for the reason that some higher frequency illumination variations are incorrectly estimated by using only a few low-frequency coefficients. As a consequence, logarithm face images should be used for recognition in the proposed approach.

Figure 5.11: Performance comparison between logarithm and non-logarithm images (Yale B Subset 4).

5.3.4 Discarding DCT Coefficients versus Discarding PCA Components

For the Eigenface method (PCA), it has been suggested that by discarding the three most significant principal components, variations due to lighting can be re-

Figure 5.12: Performance comparison between logarithm and non-logarithm images (Yale B Subset 5).

duced [18]. In [18], experimental results show that the Eigenface method performs better under variable lighting conditions after removing the first three principal components. However, the first several components not only correspond to illumination variations, but also some useful information for discrimination [18]. Besides, since the Eigenface method is highly dependent on the training samples, there is no guarantee that the first three principal components are mainly related to illumination variations. Fig. 5.14 shows the performance based on the Eigenface method by discarding different numbers of the first several principal components. It is evident that discarding first several principal components will not definitely improve the performance significantly.

5.3.5 DCT versus DFT

If the method of discarding DCT coefficients is regarded as a kind of filtering, DFT can also be employed since it is widely used as an image filtering method in the frequency domain. If DFT is employed instead of DCT, the proposed approach is similar to the so-called homomorphic filtering, an image enhancement approach which is used for contrast enhancement [91]. Homomorphic filtering can be accomplished by applying high-boost filtering or high-frequency emphasis

Figure 5.13: Performance comparison between logarithm and non-logarithm images (CMU PIE).

filtering in the logarithm domain and subsequently transforming the filtered image to the original spatial domain using inverse logarithm transform. Different from the homomorphic filtering method for image enhancement, in illumination normalization applications, low-frequency illumination variations should be completely suppressed. Hence, high-pass filtering should be employed in the logarithm domain to reduce low-frequency illumination variations.

Normally, an ideal filtering method is not recommended for image filtering because unwanted ringing behavior may be generated [91]. However, in face recognition applications, we are more concerned with recognition performance rather than image quality. Fig. 5.15, Fig. 5.16 and Fig. 5.17 show the performances based on the correlation method of high-pass filters with different transfer functions, i.e., the 2nd- and 4th-order Butterworth high-pass filters, and the ideal high-pass filter. These transfer functions achieve similar performances. The only difference is that the Butterworth high-pass filter has a smoother performance curve with varying cutoff frequencies for the reason that, unlike the ideal filter, which has a sharp edge between passed and filtered frequencies, it employs a smooth transfer function. However, the best performance is achieved using the ideal high-pass filter. Moreover, the ideal filter is computationally less complex. Consequently, ideal

Figure 5.14: Performance based on the Eigenface method by discarding different numbers of principal components (50 principal components are used, i.e., the dimension of feature vectors is 50).

filters can be used for illumination normalization for face recognition applications. Fig. 5.18 shows the filtered images using different transfer functions of high-pass filters.

Figure 5.15: Performance on the Yale B database using high-pass filters with different transfer functions (Subset 4).

The DCT is employed in this chapter because it has a few advantages over the DFT: 1) The DCT is real-valued instead of complex (i.e. it involves magnitude and phase) such that it is easier to be implemented than the DFT; 2) When the DFT transform coefficients are truncated, the Gibbs phenomenon causes the

Figure 5.16: Performance on the Yale B database using high-pass filters with different transfer functions (Subset 5).

boundary points to take on erroneous values [91]. This can be observed from Fig. 5.18; 3) The DCT is more efficient for illumination variation estimation than the DFT. This can be experimentally shown in the following part.

Similar to the definition of the energy packing efficiency (EPE) [87], we may define the variation estimation efficiency (VEE) as the performance criterion of variation estimation. For images of the same person, the VEE is the variance portion contained in the first M of N transform coefficients, i.e., for person P ,

$$VEE_P(M) = \frac{\sum_{i=0}^{M-1} E[(X_i - \bar{X}_i)^2]}{\sum_{i=0}^{N-1} E[(X_i - \bar{X}_i)^2]} \quad (5.6)$$

The average VEE of the Yale B and CMU PIE database are respectively shown in Fig. 5.19 and Fig. 5.20. Obviously, the DCT has better VEE than the DFT, especially for the first a few coefficients. In other words, DCT basis images are more correlated with illumination variations. We can also see from Table 5.3 that better recognition performance is achieved by using the DCT. (The results are the average error rates of the best performance range.)

Figure 5.17: Performance on the CMU PIE database using high-pass filters with different transfer functions.

Figure 5.18: (a) 2nd-order Butterworth filter; (b) 4th-order Butterworth filter; (c) Ideal filter. (cutoff = 0.06)

5.3.6 Performance with Mis-Aligned Face Images

The above experiments are all based on the well-aligned face images. Therefore, the higher-frequency features are well utilized for recognition. In some practical applications, face images may not be aligned well. In this section, some experimental results on slightly mis-aligned face images are presented. Since the face alignment is based on the eye coordinates, the mis-aligned face images are ob-

Table 5.3: Recognition performance comparison between DCT and DFT (Correlation)

Methods	Error rate (%)			
	Yale B			CMU PIE
	Subset 3	Subset 4	Subset 5	
DCT ($18 \leq D_{dis} \leq 25$)	0	0.18	1.71	0.36
DFT ($0.052 \leq Cutoff \leq 0.068$)	0.83	5.29	2.63	1.62

Figure 5.19: Average variation estimation efficiency (VEE) of the Yale B database (Coefficients are sorted according to the DCT discarding and DFT filtering manner).

tained by randomly adding offset errors to the eye coordinates such that there are small translation, scale and rotation variations in face images. Fig. 5.21 shows the slightly mis-aligned face images from Yale B Subset 1.

Table 5.4: Recognition performance comparison between well-aligned and mis-aligned images (Correlation, $18 \leq D_{dis} \leq 25$)

Methods	Error rate (%)			
	Yale B			CMU PIE
	Subset 3	Subset 4	Subset 5	
Well-aligned	0	0.18	1.71	0.36
Mis-aligned	3.44	6.07	5.92	28.5

As shown in Fig. 5.22, Fig. 5.23, Fig. 5.24 and Table 5.4, the overall performance on mis-aligned images is worse than that on the well aligned images. This is the major drawback of appearance-based face recognition approaches. Besides, the performance degrades earlier in terms of D_{dis} because higher-frequency features cannot be efficiently utilized. As a result, the value of D_{dis} should also be chosen by considering accuracy of the alignment procedure. As aforementioned, discarding DCT coefficients is a trade-off between low-frequency features and illumination variations. The proper D_{dis} should be chosen to minimize illumination variations as well as to keep low-frequency information as much as possible. Moreover, it

Figure 5.20: Average variation estimation efficiency (VEE) of the CMU PIE database (Coefficients are sorted according to the DCT discarding and DFT filtering manner).

Figure 5.21: Examples of mis-aligned face images from Yale B Subset 1.

also depends on how efficient the feature extraction method could utilize higher-frequency features that are essential for precise face recognition especially under large illumination variations. Nevertheless, the experimental results on these two databases show that recognition performance can be significantly improved when $D_{dis} > 5$ even with the face images cannot be well aligned.

5.4 Conclusions

A novel illumination normalization approach is proposed in this chapter. Illumination variations under different lighting conditions can be significantly reduced by discarding low-frequency DCT coefficients in the logarithm domain. Our approach has several advantages: 1) No modelling step and bootstrap sets are required; 2) Our approach is very fast and it can be easily implemented in a real-time face

Figure 5.22: Performance on the mis-aligned Yale B database with different D_{dis} (Correlation).

recognition system; 3) The proposed approach outperforms most of existing approaches. Nevertheless, the shadowing and specularity problems are not perfectly solved because they lie in the same frequency band as some facial features. Furthermore, higher frequency facial features are more difficult to extract while poses and expressions change.

Figure 5.23: Performance on the mis-aligned Yale B database with different D_{dis} (Eigenfaces).

Figure 5.24: Performance on the mis-aligned CMU PIE database with different D_{dis} .

Chapter 6

Higher-Order Invariant Representation for Face Recognition

A higher-order invariant representation for face images is proposed to achieve translation, scale and rotation invariance. Thus, the freedom of the image space is largely reduced to obtain more generalization ability and the face alignment procedure, which is a non-trivial and unreliable task, is not required any more. Gradient images are employed to extract salient facial features and the extracted binary, gradient or gray-level feature images can be used for obtaining higher-order representations.

6.1 Introduction

Appearance-based face recognition approaches have attracted a great deal of attentions because of their easy implementation and low computational cost. However, one of the major limitations of appearance-based approaches is that they are sensitive to geometric transformations in translation, scale and in-plane rota-

tion. In practice, there are a lot of geometric variations in face images. So far, face detection techniques cannot perfectly locate face images and determine the exact rotation angles. Therefore, some facial features, especially higher-frequency features, cannot be efficiently utilized [6, 7].

In order to make good use of these useful facial features so as to precisely recognize face images, a large number of training samples with possible variations are needed, which seems impractical. There are two ways to deal with this problem: (1) Alignment and normalization; (2) Invariant representations. Face alignment and normalization are usually based on detection of important facial features, e.g. the mostly used eye detection technique [118, 119, 120]. Provided the coordinates of eyes are used as reference points, face images with the same pose can be normalized to the same scale and orientation. However, accurate and reliable detection of facial features is a non-trivial task since there are a lot of variations in face images such as poses and illumination. Besides, a computationally expensive search process is necessary for detecting facial features. Another way is to employ invariant pattern recognition techniques of [121]. Invariant representations can be used to improve the generalization of features as well as reduce the complexity of the image space. A few invariant approaches have been proposed to handle geometric changes of images such as Fourier-Mellin Transform [122], Triple-correlation [123], geometrical moments [124], Zernike moments [125], etc. In [126], the face representation called spectroface which combines the wavelet transform and holist Fourier invariant features is proposed to achieve geometric invariance. Zernike moments are also employed for face recognition in [127] to achieve better generalization. One drawback of the aforementioned invariant representations of moments is that they are sensitive to noise because the invariance relies on accurate determination of the centroid. The situation will be even worse if there are illumination changes in images.

Higher-order neural networks (HONN) have also been applied in invariant pattern recognition [70, 71, 72, 73, 74, 75, 76]. The most important advantage of HONN

over conventional neural networks is that invariance can be built directly into the architecture of HONN. It has been shown that third-order neural networks is sufficient to achieve translation, scale and rotation invariant recognition. In [74], experimental results on digit recognition show that third-order networks are superior to the use of the Zernike moments followed by a conventional neural network classifier except patterns corrupted by noise. The improved higher-order neural networks can also be applied in distortion invariant pattern recognition [128, 129]. Another advantage of HONN is that they can be trained rapidly [72]. Nevertheless, the most severe problem of HONN is that a great deal of computational space is required. For the third-order network, the number of possible triplet combinations increases as the size of the input field increases. Coarse-coding representation has been used to overcome this limitation for binary images [75]. High resolution images are represented using a set of low resolution images such that the input fields for each image are largely reduced. It is difficult to apply HONN for face recognition for the reason that face images are normally high dimensional and a lot of useful facial features exist in gray-level information. In [77], a third-order neural network was employed for face recognition. Since HONN is usually used for binary images, gray-level face images are represented using a set of quantized isodensity binary images and thus the coarse coding algorithm can be employed for isodensity images to reduce memory requirements and computational complexity. The major drawback in using isodensity information is that it is very sensitive to lighting variations on face images although the performance can be somewhat improved using appropriate preprocessing methods.

In this chapter, we attempt to directly use gray-level information of face images in third-order neural networks as well as to reduce the number of input fields while keeping the most important facial features. Moreover, it has been shown that third-order neural networks are a specific category of triple-correlation-based neural networks [123]. Based on the advantages of conventional neural networks as a classifier, face images can be expressed by triple-correlation invariant representations which can be further fed into conventional neural network classifiers to

achieve better recognition performance.

The remainder of this chapter is organized as follows. In Section 6.2, we describe the preprocessing and feature extraction step. A third-order neural networks is introduced in Section 6.3. In Section 6.4, triple-correlation-based invariant representation is described. Experimental results and discussions are presented in Section 6.5. Finally, conclusions are drawn in Section 6.6.

6.2 Preprocessing and Feature Extraction

In order to obtain the invariant representation, the preprocessing step is necessary. In some literature, examples of geometric variations are just obtained by translating, rotating and scaling the whole image (See Fig. 6.1). However, in practical face recognition applications, cropping of face images is different mainly due to changes in rotation because face detection techniques cannot determine geometric variations especially the rotation (face size and position can only be approximately determined). Fig. 6.2 shows two close-cropped images with geometric variations. In this chapter, only close-cropped face images excluding hair and face contour are discussed because inner local facial features such as eyes, nose and mouth are more stable. For the face images with practical geometric variations, the conventional way of dealing with such variations may fail because they have boundaries with non-zero values. To eliminate the cropping effect, salient local facial features should be extracted and used for recognition and others are regarded as zero-valued background. It should be noted that in order to achieve geometric invariance while using the entire face image including hair and contour, face regions should also be extracted and the background should be set to zero values.

Figure 6.1: Example of face images with ideal geometric variations

Figure 6.2: Example of face images with practical geometric variations

6.2.1 Gradient Image

To our knowledge, a lot of discriminating facial features are high-frequency information. In [40], face images are first preprocessed to edge binary images in which mainly high-frequency information is preserved. Their method based on the modified Hausdorff distance achieves good recognition performance. As demonstrated in Chapter 5, high-frequency features are less sensitive to illumination variations whereas these features are difficult to be extracted especially for mis-aligned face images. Since the edge map approach is a binary template matching approach, images should be well aligned. In this chapter, higher-order invariant representations are used to achieve good generalization and alignment is not required any more. Therefore, edge images can be used for recognition. Furthermore, the extracted gradient images can also be used since they contain more information than the binary edge images. The gradient of an image can be obtained based on the

partial derivatives of two directions at every pixel location. The derivatives can be implemented by using the special masks such as Roberts cross-gradient, Prewitt and Sobel [91]. In this chapter, the Sobel operator is employed because it possesses noise-suppression characteristics. Fig. 6.3 shows the Sobel gradient image and the thresholded binary image (20% highest gradient values). As shown in Fig. 6.3(b), the main facial features, such as eyes, eye brown, nose and mouth, can be extracted because large gradient values are around those facial features. The cropping boundary problem can be solved by setting regions of low gradient value to zero. At the same time, the the input fields to the HONN are largely reduced.

Figure 6.3: Feature extraction based on Sobel operator: (a) Original image; (b) Sobel gradient image; (c) Thresholded binary image.

6.2.2 Illumination and Noise Reduction

There may be some illumination variations and noise in face images which will result in unreliable feature extraction and degrade subsequent recognition performance. As described in Chapter 5, logarithm transform can be used to reduce illumination variations. Fig. 6.4 shows the logarithm face image, its Sobel gradient image [91] and the thresholded binary image. Obviously, compared with the binary image of Fig. 6.3(c), there is less noise in the binary image obtained from the logarithm image because illumination variations such as specularities are reduced by using the logarithm transform. In order to remove noise from the binary image, morphological operations can be used. Filters constructed from morpho-

logical operations such as opening and closing are efficient to remove noise while distorting the image as little as possible [91]. Here, closing followed by opening morphological operations are employed to fill discontinuities of the extracted regions and then eliminate pepper noise. Fig. 6.5(a) shows the binary image after these operations. Fig. 6.5(b) and Fig. 6.5(c) show the extracted Sobel gradient image and the extracted logarithm gray-level image respectively. Moreover, the partially included contour of the face should be removed to minimize differences caused by cropping. Elliptical or circular cropping may be adopted in order to efficiently crop out the contour of a face. The high gradient values near the cropping boundary caused by zero-padding should also be removed.

Figure 6.4: Feature extraction on logarithm image: (a) Logarithm image; (b) Sobel gradient image; (c) Thresholded binary image.

Figure 6.5: Extracted feature images after morphological operations: (a) Binary image after morphological operations; (b) Extracted Sobel gradient image; (c) Extracted logarithm gray-level image.

6.3 Third-Order Neural Networks

Figure 6.6: Third-order neural network with 4 inputs and 1 output.

Third-order neural networks can be used to build simultaneous invariance to translation, scale, and in-plane rotation. Fig. 6.6 shows a third-order neural network with 4 inputs and 1 output. All the inputs to a third-order neural network are first combined in triplets and then the output is determined from the weighted sum of these products. The output of the i th node denoted by y_i can be represented by

$$y_i = f\left(\sum_j \sum_k \sum_l x_j x_k x_l\right) \quad (6.1)$$

where x_j is the excitation value of the j th input node, and w_{ijkl} is the weight connecting the product of j th, k th and l th inputs to the i th output node, and f is the nonlinear activation function (the threshold function is normally used). Fig. 6.7 shows an example of extracted binary face images with translation, scale and rotation variations. (It should be noted that the practical situation has been shown in Fig. 6.2. Fig. 6.7 is simply for the convenience of showing the properties of third-order neural networks). As shown in this figure, the triplets of input pixels define triangles with included angles (a, b, c) which are invariant under these three

types of geometric transformations. Therefore, invariance can be achieved using a third-order network in which the weights are constrained such that all combinations of three pixels defining similar triangles are connected to the output with the same weight. This can be written as

$$w_{ijkl} = w_{iabc} = w_{icab} = w_{bca} \tag{6.2}$$

From (6.2), we can see that similar triangles in the same direction share the same weight. It should be noted that if there are no equal included angles, these three angles (a, b, c) can be listed in clockwise or counter-clockwise directions, i.e. $w_{iabc} \neq w_{ibac}$. If all similar triangles share the same weight regardless of the listing direction, it is also invariant against reflection transforms of images.

Figure 6.7: Built-in invariance of face images with translation, scale and rotation variations

Face images are usually high dimensional data. There are still a large number of possible combinations of triplets even after the preprocessing procedure. Nevertheless, the number of weights can be reduced by assigning approximately similar triangles to the same equivalence class [74]. In fact, the number of weights is determined by the quantization of included angles of triangles. By quantizing the angles, not only the number of weights is largely reduced, but also small local

distortion robustness and noise insensitivity are enhanced. In order to keep sufficient discriminating features as well as to reduce the number of weights, proper quantization degree should be chosen.

The third-order neural networks can be trained using a simple perceptron-like rule of the form as follows [75]:

$$\Delta w_{ijkl} = (t_i - y_i)x_jx_kx_l \quad (6.3)$$

where t_i is the expected output of unit i , y_i is the actual output. x_j , x_k and x_l are the inputs.

6.4 Higher-Order Invariant Representation

In [123], it has been proven that third-order neural networks are a specific category of triple-correlation-based neural networks. A third-order neural network is equivalent to the special case invariant representation which is obtained from the zeroth frequency coefficient of the Fourier transform of triple correlation. In this chapter, we employ this invariant representation for face recognition. It has been proven that there is a one-to-one relation between a signal and its triple correlation [130]. Invariant representation is obtained by sacrificing the uniqueness of the triple correlation. It has also been experimentally shown that such representation contains sufficient features for pattern recognition as well as achieves invariance to geometric transformations. The invariant representation can be regarded as transforming the original image space to a 2-D (a, b) image space (a, b, c are included angles of possible similar triangles and $a \leq b \leq c$). This is depicted in the (a, b) space as illustrated in Fig. 6.8. Similarly, the dimensionality (i.e. the number of weights) of the representation image is determined by quantization of the included angles of triangles. The representation space can be separated into two parts if the order of listing included angles, i.e. clockwise or counter-clockwise, is not ignored. To achieve invariance to reflection transformation, only the half part is required.

Figure 6.8: Invariant image space with respect to angles a and b .

For the invariant representation, conventional dimensionality reduction or feature selection approaches such as PCA, LDA and ICA can further be employed to obtain more representative and discriminating features. Since the human face is a non-rigid object, there still exist a lot of non-linear variations such as expressions and poses. Efficient nonlinear classifiers, e.g. conventional multi-layer neural networks, can be employed to achieve better performance although it has been shown in [73] that a third-order neural network can provide nonlinear separation using only a single layer. If the triple-correlation-based invariant representations are directly fed into the neural network, it can be regarded as a hybrid neural network combining the third-order neural network and the conventional neural network.

6.4.1 Partial Connectivity Strategies

After the feature extraction step, the computational cost is largely reduced. However, the computational burden is still heavy since face images are normally high dimensional. In order to further reduce the computational complexity, the partial connectivity strategies can be employed [76]. Some triplets can be ignored if enough discriminating features can be represented by others. The key idea of partial connectivity strategies is that only triplets whose inter-pixel distances are less (or more) than the predefined bound are formed. In this chapter, triplets with short inter-pixel distances are ignored. As shown in Fig 6.9, the triplet formed by the dashed line may be ignored if the shortest side of the triangle is less than the

predefined distance bound.

Figure 6.9: Example of partial connectivity.

6.5 Experimental Results

In this chapter, the Yale Face Database B is used to evaluate the proposed method for invariant face recognition. We simply test the generalization ability of invariant representations with respect to geometric variations such as translation, scale and rotation. As a consequence, only frontal face images without expression variations are used here. As shown in Fig. 6.10, only one face image per person from Subset 1 are used as training samples. There are random geometric variations in the testing images. In a practical situation, the most serious geometric variation is rotation since the position and scale can be approximately determined by using face detection techniques. The face images are all resized to 80×70 and 20 % pixels are extracted to reduce the computational complexity. In the following experiments, Linear Least Square (LLS) is used as a linear classifier. The triple-correlation-based representations are all normalized to zero mean and unit variance to improve the robustness against scale and illumination variations. We define the partial connectivity distance bound as D_b (in pixel unit) and quantization degree as α . As described in Section 6.4, there are two parts of the higher-order invariant representation if the reflection of images is not ignored. However, our experiments

show that the two kinds of representations achieve similar performance. Therefore, we use representations ignoring the order of listing included angles to achieve reflection insensitiveness and computational simplicity.

Figure 6.10: Example of face images used for training and testing: (a) Training; (b) Testing (Subset 1).

6.5.1 Original Image versus Logarithm Image

As shown in Section 6.2.2, there is less noise in the extracted binary image. The results in Table 6.1 also show that feature extraction based on logarithm images outperforms the one based on original images. Table 6.1 also shows that the performance on the extracted Sobel gradient image is better than the extracted binary images and gray-level images. Compared with binary images, the gradient images possess more information of facial features and they are less sensitive to the aforementioned thresholding procedure because the abandoned pixels are usually small values compared with the ones that are retained. The gray-level images are also more sensitive to thresholding although they contain even more information than gradient images. As a consequence, the extracted Sobel gradient images are more

suitable as facial features for subsequent higher-order invariant representations.

Table 6.1: Feature extraction comparisons between original image and logarithm image ($D_b = 20$, $\alpha = 0.5^\circ$).

	Error rate (%)		
	Binary image	Gradient image	Gray-level image
Original	23.3	20.0	35.0
Logarithm	21.7	11.7	36.7

Table 6.2: Results of the eigenface approach on face images with geometric variations.

	Error rate (%)
Original	13.3
Gradient image	28.3

Table 6.2 shows the results of the eigenface approach on original images and extracted gradient images. It is difficult for conventional appearance-based approaches to recognize faces using gradient images with geometric variations because pixels corresponding to the same facial feature cannot be well matched. Instead, higher-order invariant representations ignore geometric variations of pixels such that higher frequency features such as gradient information can be utilized. Fig. 6.11(a) and Fig. 6.11(c) are two extracted gradient images with geometric variations and their corresponding higher-order invariant representation images are shown in Fig. 6.11(b) and Fig. 6.11(d) respectively. It should be noted that logarithm invariant representation images are illustrated instead of original ones for better displaying. It is evident that geometric variations are ignored in the invariant representations. Nevertheless, they are not perfectly identical because there are other variations in the gradient images inherited from the feature extraction step.

6.5.2 Effect of Partial Connectivity Strategy

As described in Section 6.4.1, the partial connectivity strategy could be employed to reduce computational cost. Different from [76], in this chapter, only triplets

Figure 6.11: Higher-order invariant representations of gradient images with translation, scale and rotation variations ($D_b = 20$, $\alpha = 0.5^\circ$).

whose inter-pixel distances are more than the predefined bound D_b are formed. As illustrated in Fig. 6.12, the best performance is achieved when the partial connectivity distance bound D_b is around the value of 20. Small D_b does not necessarily improve the performance because triplets with short distances are more sensitive to the instability in the feature extraction step and other variations of face images.

Figure 6.12: Performance with different partial connectivity distance bound D_b ($\alpha = 0.5$).

6.5.3 Performance with Different Quantization Degrees of Included Angles

The dimensionality of higher-order invariant representations are determined by the quantization procedure. Proper quantization degree α should be chosen to obtain enough discriminating features as well as robustness against image distortion and noise. Higher-order invariant representations of Fig. 6.11(a) corresponding to different quantization angles are shown in Fig. 6.13. It is clear that the smaller the quantization angles are, the higher resolution the representation images obtain. Table 6.3 shows the number of values in representations corresponding to different quantization degrees. Fig. 6.14 shows the performance with different quantization degrees. The best performance is achieved when α is around the value of 0.3° . Smaller quantization degree means more information obtained whereas the storage requirement will increase significantly. Moreover, as shown in Fig. 6.14, smaller quantization degree not necessarily improves the recognition performance.

Figure 6.13: Higher-order invariant representations with different quantization degrees ($D_b \geq 20$).

6.5.4 Performance with DCT Illumination Normalization in Logarithm Domain

As shown in Chapter 5, discarding DCT coefficients in logarithm domain greatly improves the recognition performance of face images with large illumination variations. However, the improved performance will degrade if face images are mis-aligned because higher-frequency features cannot be utilized using conventional appearance-based approaches. As a consequence, feature extraction methods insensitive to mis-alignment or geometric variations are desirable. In this section, experimental results on mis-aligned images with illumination variations are pre-

Table 6.3: Number of values in representations with different α .

$\alpha(^{\circ})$	2	1	0.5	0.3	0.1
No. of values	490	1819	6953	18537	117535

Figure 6.14: Performance with different quantization degree α ($D_b = 20$).

sented (See Fig. 6.15).

Figure 6.15: Example of mis-aligned face images with illumination variations from Yale B Subset 3.

Fig. 6.16 shows the performance of higher-order representations with different D_{dis} (refer to Chapter 5). Even for mis-aligned images, the recognition performance can be improved and the value of D_{dis} for the best result is close to the one based on well-aligned face images in Chapter 5 (i.e. around 20). However, the performance on face images with large illumination variations is still worse because the feature

extraction step is very sensitive to illumination variations. As shown in Fig. 6.17, the major drawback of using gradient images for feature extraction is that higher-frequency illumination variations such as shadowing cannot be ignored. Gradient images are not suitable for recognizing face images with large shadowing variations.

Figure 6.16: Performance with different D_{dis} based on higher-order representations.

Figure 6.17: Example of a face image with shadowing variations and the extracted gradient feature image.

6.5.5 Performance on the ORL Face Database

The invariant representations are simply tested on the frontal images of the Yale Face Database B in order to eliminate other factors which may affect the recognition performance. The main goal of this chapter is to demonstrate the implemen-

tation possibility and the geometric invariance ability of the higher-order invariant representations based on gradient face images. In order to show the performance on larger face database with other variations, experimental results on the ORL face database are shown in Table 6.4. Like the evaluation method in Chapter 4, 5 images are randomly selected as the training samples and another 5 images as test images. The average error rate is calculated based on 10 runs. Although the performance is worse than other approaches listed in Chapter 4, it should be noted that only partial gradient face images, i.e. mainly higher-frequency features, are used here.

Table 6.4: Performance on the ORL face database ($\alpha = 0.5$, $D_b = 20$).

$E_{min}(\%)$	$E_{max}(\%)$	$E_{ave}(\%)$	$\sigma_E(\%)$
7.5	10	9.05	1.25

6.6 Conclusions

In this chapter, a face recognition approach based on higher-order invariant representations is proposed. First, gradient images are employed to extract salient facial features and then the extracted binary, gradient or gray-level feature images can be used for obtaining higher-order representations. The feature extraction step reduces the computational complexity as well as eliminates the boundary differences caused by cropping. Next, higher-order representations are computed using partial connectivity strategies whereby the computational time is further reduced and insensitiveness to instability in the feature extraction step is improved. Experiments on face images with translation, scale and rotation geometric variations show that higher-order representations achieve better performance than conventional appearance-based approaches. In particular, performance on mis-aligned face images normalized by discarding DCT coefficients in logarithm domain can be improved because higher frequency features may be utilized by using higher-order representations. Nevertheless, there are a few limitations in the higher-order

invariant representation: (1) It is still too computationally expensive for real-time face recognition; (2) Performance greatly relies on the feature extraction step which is unreliable. Therefore, computational complexity reduction and feature extraction and selection for higher-order representations should be explored in future work.

Chapter 7

Conclusions and Recommendations for Future Work

7.1 Conclusions

In this thesis, a few efficient approaches for some challenging problems in face recognition have been proposed and discussed. It has been proven that the well-known PCA and LDA methods can be directly implemented in the DCT or block DCT domain and the projection results are exactly the same as the one obtained from the spatial domain. Therefore, DCT coefficients of JPEG images can be directly used to reduce storage requirements and by initially discarding an appropriate number of DCT coefficients with less information, computational complexity of PCA and LDA can be significantly reduced. Experiments on the FERET database show that the results on original images and on truncated and quantized JPEG coefficients are close. Moreover, for the D-LDA method, performance on DCT coefficients is better than on original images.

In order to achieve real-time face recognition, a high-speed face recognition ap-

proach based on the DCT, LDA and RBF neural networks has been proposed. The low-frequency DCT coefficients are first used as basic facial features and after a proposed clustering step, the LDA is implemented in the truncated DCT domain to further extract discriminating and invariant features. The DCT+LDA scheme is more computationally efficient than the Fisherface approach. Normally, face images are distributed in a nonlinear feature space because there are a lot of variations in faces. RBF neural networks incorporated with the proposed clustering and parameter selection method can achieve efficient nonlinear classification as well as high-speed training and recognition. The low-frequency DCT coefficient features are less sensitive to alignment, distortions and pose variations. However, some discriminating and detailed facial features lie in higher-frequency domain such that this approach cannot be guaranteed to recognize faces with similar looks. Nevertheless, for precise face recognition systems, it still can be employed as a coarse recognition step to make high speed recognition possible.

In this thesis, based on the reflectance theory, it has been proven that discarding low-frequency DCT coefficients in the logarithm domain can efficiently reduce illumination variations on faces. Experimental results on the Yale B database and CMU PIE database show that the proposed approach improves the performance significantly for face images with varying lighting conditions. Compared with existing approaches, our approach achieves comparable results while requires less computational cost because no modelling steps are involved. In addition, experiments demonstrate that good recognition performance can be achieved by using only higher-frequency features if face images are well-aligned. This inspires us to explore face recognition approaches which can utilize higher-frequency features. If higher-frequency information could be utilized well, precise recognition based on detailed features becomes possible. One of the shortcomings of this approach is that some low-frequency features will be removed together with illumination variations although these features may be impaired by illumination variations. Besides, high-frequency illumination variations, e.g. shadowing, cannot be removed.

For the appearance-based face recognition approaches, higher frequency facial features are difficult to be utilized adequately if face images are not aligned well. Normally, face detection techniques only approximately locate the face and there are geometric variations, especially changes in rotation. A higher-order invariant representation for a face image has been investigated in this thesis. Gradient images are employed to extract salient facial features and then used for obtaining higher-order representations. The feature extraction step is essential to cut down computational cost as well as to eliminate boundary differences caused by cropping. Experiments on face images with practical geometric variations show that higher-order representations achieve better performance than the conventional appearance-based approaches. Combined with the proposed illumination normalization approach based on discarding DCT coefficients, robustness against illumination variations can also be improved. One of its limitations is that it is computationally expensive such that it cannot be applied in real-time systems. Nevertheless, unlike some other feature matching approaches [50, 52, 53], the computational complexity of the invariant representation will not increase with the number of face images in a face database. Moreover, the feature extraction step is not reliable especially for images with large illumination variations. Therefore, desirable performances cannot be achieved.

7.2 Recommendations for Future Work

Future work could be focused on the following topics:

- (1) As aforementioned, higher frequency facial features are more robust against changes in illumination and a lot of important shape information lies in higher frequency features. It is more desirable if similar faces are need to be identified correctly. However, they are not easily extracted and utilized due to mis-alignment, pose and expression variations. For most of appearance-based approaches, it is a difficult task to obtain these higher frequency fea-

tures. Distortion invariant matching approaches may be implemented on the normalized face images using the idea of discarding DCT such that higher frequency features can be extracted efficiently and achieve illumination insensitiveness. For example, the distortion matching approaches such as DLA and EBGM can be applied on the normalized face images with higher frequency features. However, these distortion matching approaches are not expected to deal with very large variations. From the point of my view, to achieve better performance, we need to obtain training samples for large variations as many as possible because it is difficult to match two faces with great differences even for distortion invariant matching approaches. Moreover, classifiers with good generalization ability are required to deal with faces with large variations.

- (2) The recognition performance based on higher-order invariant representations are not so ideal as expected. Higher-order invariant features should be further investigated. It is desirable to extract discriminating and invariant features directly from the representations to improve subsequent recognition performance. In fact, the higher-order representations can be regarded as images with their own specific properties in a geometric invariance space. Based on appropriate analysis on these properties, conventional linear or nonlinear subspace learning methods, e.g. PCA, ICA, KPCA and GDA, can be further employed in such space. Hopefully, the recognition performance can be improved especially for face images with other variations. It should be noted that the higher-order representations are global features. It would be more promising if local features could be extracted by using higher-order invariant representations.
- (3) Efficient matching or similarity measurement based on local features should be explored. Holistic approaches are sensitive to all sorts of variations because the similarity between faces is measured based on the entire information in which some facial features may be varied or ruined. Local facial

features can be employed for recognition regardless of the varied and ruined features since some unique local facial features may be sufficient to identify the subject. For face recognition, global similarity measurement is not as ideal as expected because human faces are non-rigid objects and there are a variety of variations. Some efficient similarity measurement methods based on partial information should be investigated instead of conventional metric distance measurement. For example, similarity between faces could be measured based on the scores that how many local features are matched. This kind of measurement is more robust against some illumination variations, occlusions etc. It can also be employed in distortion invariant matching approaches. With the partial similarity measurement, better recognition performance under variable circumstances is expected to be achieved.

- (4) Online training could be investigated. In face recognition applications, the number of training samples is directly related to the performance. As shown in Chapter 4, the performance becomes better as the number of training samples increases. More training samples there are, more generalization ability a face recognition system can achieve. On the other hand, online training makes it possible to keep the training database updated because human faces change with time. For some applications, it is possible to obtain more training samples while the face recognition system is running. For example, face images captured in running time with very high score can be put into the training sample database. Furthermore, for dynamic matching applications, if a face can be tracked, many face images with pose and expression variations can be obtained as training samples as long as one of the face images can be identified.

Author's Publications

- (1) W. Chen, M. J. Er and S. Wu, "Face Recognition Using Discrete Cosine Transform and RBF Neural Networks", Chapter in *Intelligent Multimedia Processing with Soft Computing* edited by Y. P. Tang, K. H. Yam and L. Wang, Springer, 2005.
- (2) W. Chen, M. J. Er and S. Wu, "PCA and LDA in DCT domain", *Pattern Recognition Letters*, Vol. 26, pp. 2474-2482, 2005.
- (3) M. J. Er, S. Wu and W. Chen, "Face Recognition Using a RBF Neural Classifier with Hybrid Learning", Chapter in *Intelligent Sensory Evaluation* edited by D. Ruan and X. Zeng, Springer, 2004.
- (4) M. J. Er, W. Chen and S. Wu, "High-speed Face Recognition Using Discrete Cosine Transform and RBF Neural Networks", *IEEE Trans. on Neural Networks*, Vol. 16, No. 3, pp. 679-691, May, 2005.
- (5) W. Chen, M. J. Er and S. Wu, "Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithm Domain", *IEEE Trans. on Systems, Man, and Cybernetics-Part B*. Vol. 36, No. 2, pp. 458-466, April, 2006.
- (6) W. Chen, M. J. Er and S. Wu, "Illumination Compensation and Normalization Using Logarithm and Discrete Cosine Transform", in *Proceeding of IEEE International Conference on Control, Automation, Robotics and Vision*, Kunming, China. December 6-9, 2004.

-
- (7) W. Chen, M. J. Er and S. Wu, "Fast PCA and LDA for JPEG Images", in *Proceeding of International Conference on Computer Recognition Systems CORES'05*, Wroclaw, Poland. pp. 137-144, 2005.
- (8) W. Chen, M. J. Er and S. Wu, "Face recognition based on higher-order invariant representations", submitted to *Pattern Recognition*.

Bibliography

- [1] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey", *Proceedings of the IEEE*, vol. 83, pp. 705–740, 1995.
- [2] C. Kotropoulos and I. Pitas, "Face authentication based on morphological grid matching ", in *Proceedings of International Conference on Image Processing*, 1997, vol. 1, pp. 105–108.
- [3] K. Jonsson, J. Kittler, Y.P. Li, and J. Matas, "Support vector machine for face authentication", *Image and Vision Computing*, vol. 20, pp. 369–375, 2002.
- [4] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face recognition algorithms", *Image and Vision Computing*, vol. 16, no. 5, pp. 295–306, 1998.
- [5] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [6] B. K. L. Erik Hjelmås, "Face detection: A survey", *Computer Vision and Image Understanding*, vol. 3, no. 3, pp. 236–274, 2001.
- [7] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.

- [8] S. Haykin, *Neural Networks, A Comprehensive Foundation*, New York: Macmillan, 1994.
- [9] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press Inc, New York, 1995.
- [10] S. Z. Li and J. Lu, "Face recognition using the nearest feature line method", *IEEE Transactions on Neural Networks*, vol. 10, no. 2, pp. 439–443, 1999.
- [11] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 4, pp. 449–465, 2004.
- [12] M. Kirby and L. Sirvoich, "Application of the Karhunen-Loeve procedure for the characterization of human faces", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 103–108, 1990.
- [13] M. A. Turk and A. P. Pentland, "Eigen faces for recognition", *Journal of cognitive Neuroscience*, vol. 3, pp. 71–86, 1991.
- [14] S. Cagnoni, A. Poggi, and G. Porcari, "A modified modular eigenspace approach to face recognition", in *Proceedings of The 10th International Conference on Image Analysis and Processing*, Venice, Italy, 1999, pp. 490–495.
- [15] H.-C. Kim, D. Kim, and S. Y. Bang, "Face recognition using the mixture-of-eigenfaces method", *Pattern Recognition Letters*, vol. 23, pp. 1549–1558, 2002.
- [16] V. Brennan and J. Principe, "Face classification using a multiresolution principal component analysis", in *Proceedings of IEEE Workshop on Neural Networks for Signal Processing*, 1998, pp. 506–515.
- [17] P. S. Penev and J. J. Atick, "Local feature analysis: A general statistical theory for object representation", *Network: Computation in Neural Systems*, vol. 7, no. 3.

- [18] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces versus fisherfaces: Recognition using class specific linear projection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [19] A. Martinez and A. Kak, "PCA versus LDA", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.
- [20] Chen L.-F., H.-Y. M. Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu, "A new LDA-based face recognition system which can solve the small sample size problem", *Pattern Recognition*, vol. 33, pp. 1713–1726, 2000.
- [21] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data - with application to face recognition", *Pattern Recognition*, vol. 34, pp. 2067–2070, 2001.
- [22] Y.-F. Guo, S.-J. Li, J.-Y. Yang, T.-T. Shu, and L.-D. Wu, "A generalized Foley-Sammon transform based on generalized Fisher discriminant criterion and its application to face recognition", *Pattern Recognition Letters*, vol. 24, pp. 147–158, 2003.
- [23] R. Lotlikar and R. Kothari, "Fractional-step dimensionality reduction", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 2, pp. 623–627, 2000.
- [24] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using LDA-based algorithms", *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 195–200, 2003.
- [25] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis", *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1450–1464, 2002.

- [26] M. S. Bartlett, H. M. Lades, and T. J. Sejnowski, "Independent component representations for face recognition", in *Proceedings of the SPIE, Conference on Human Vision and Electronic Imaging III*, 1998, pp. 528–539.
- [27] P. C. Yuen and J. H. Lai, "Face representation using independent component analysis", *Pattern Recognition*, vol. 35, pp. 1247–1257, 2002.
- [28] B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridgea, "Recognizing faces with PCA and ICA", *Computer Vision and Image Understanding*, vol. 91, pp. 115–137, 2003.
- [29] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution", *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [30] A. Hyvärinen, "The fixed-point algorithm and maximum likelihood estimation for independent component analysis", *Neural Processing Letters*, vol. 10, pp. 1–5, 1999.
- [31] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction", *Science*, vol. 290, pp. 2319–2323, 2000.
- [32] M.-H. Yang, "Face recognition using extended isomap", in *Proceedings of International Conference on Image Processing*, New York, 2002.
- [33] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding", *Science*, vol. 290, pp. 2323–2326, 2000.
- [34] B. Schölkopf, A. Smola, and K. Müller, "Nonlinear component analysis as a kernel eigenvalue problem", *Neural Computation*, vol. 10, no. 5, pp. 2299–1319, 1998.
- [35] M.-H. Yang, N. Ahuja, and D. J. Kriegman, "Face recognition using kernel eigenfaces", in *Proceedings of International Conference on Image Processing*, 1998, pp. 506–515.

- [36] G. Baudat and F. Anouar, “Generalized discriminant analysis using a kernel approach”, *Neural Computation*, vol. 12, pp. 2385–2404, 2000.
- [37] M.-H. Yang, “Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods”, in *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, Washington D. C., USA, 2002, pp. 215–220.
- [38] T. Kanade, *The Nature of Statistical Learning Theory*, New York: Springer-Verlag, 1995.
- [39] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, “Face recognition using kernel direct discriminant analysis algorithms”, *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 117–126, 2003.
- [40] B. Takács, “Comparing face images using the modified Hausdorff distance”, *Pattern Recognition*, vol. 31, no. 12, pp. 1873–1881, 1998.
- [41] Y. Gao and M. K. H. Leung, “Face recognition using line edge map”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 764–779, 2002.
- [42] F. Samaria, *Face Recognition Using Hidden Markov Models*, PhD thesis, Trinity College, University of Cambridge, Cambridge, 1994.
- [43] B. Achermann and H. Bunke, “Combination of face classifier for person identification”, in *Proceedings of International Conference Pattern Recognition*, Vienna, 1996, pp. 416–420.
- [44] S. Eickeler, S. Müller, and G. Rigoll, “High quality face recognition in JPEG compressed images”, in *Proceedings of International Conference on Image Processing*, Kobe, Japan, 1999.
- [45] V. V. Kohir and U. B. Desai, “Face recognition using a DCT-HMM approach”, in *Proceedings of IEEE Workshop on Application of Computer Vision*, Princeton, USA, 1998.

- [46] L. E. Baum, "An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes", *Inequalities*, pp. 1–8, 1972.
- [47] G. D. Forney, "The Viterbi algorithm", *Proceedings of the IEEE*, vol. 61, pp. 268–278, 1973.
- [48] T. Kanade, *Picture Processing by Computer Complex and Recognition of Human Faces*, PhD thesis, Kyoto University, 1973.
- [49] R. Brunelli and T. Poggio, "Face recognition: Features versus templates", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1042–1053, 1993.
- [50] M. Lades, J. C. Vorbrüggen, J. Buhmann, Jörg Lange, C. Malsburg, R. P. Würtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 3, pp. 300–310, 1993.
- [51] J. G. Daugman, "Complete discrete 2d gabor transform by neural networks for image analysis and compression", *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. 36, pp. 1169–1179, 1988.
- [52] L. Wiskott and C. von der Malsburg, "Face recognition by dynamic link matching", *Lateral Interactions in the Cortex: Structure and Function*, Chapter 11, 1996.
- [53] L. Wiskott, J.M. Fellous, N. Krüger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.
- [54] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic interpretation and coding of face images using flexible models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 743–756, 1997.

- [55] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models – their training and application", *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
- [56] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.
- [57] T. F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models", in *Proceedings of the European Conference on Computer Vision*, 1998, vol. 2, pp. 484–498.
- [58] G. J. Edwards, T. F. Cootes, and C. J. Taylor, "Face recognition using active appearance models", in *Proceedings of the European Conference on Computer Vision*, 1998, vol. 2, pp. 581–695.
- [59] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach", *IEEE Transactions on Neural Networks, Special Issue on Neural Networks and Pattern Recognition*, vol. 8, no. 1, pp. 98–113, 1997.
- [60] Y. Le Cun, B. Boser, J. S. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, "Handwritten digit recognition with a backpropagation neural network", in *Advances in Neural Information Processing Systems 2*, 1990, pp. 396–404.
- [61] S.-H. Lin, S.-Y. Kung, and L.-J. Lin, "Face recognition/detection by probabilistic decision-based neural network", *IEEE Transactions on Neural Networks*, vol. 8, pp. 114–132, 1997.
- [62] J. Mao and A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection", *IEEE Transactions on Neural Networks*, vol. 6, pp. 296–317, 1995.

- [63] D. Valentin, H. Abdi, A. J. O'Toole, and G. W. Cottrell, "Connectionist models of face processing: a survey", *Pattern Recognition*, vol. 27, pp. 1209–1230, 1994.
- [64] J. L. Yuan and T. L. Fine, "Neural-network design for small training sets of high dimension", *IEEE Transactions on Neural Networks*, vol. 9, pp. 266–280, 1998.
- [65] M. J. Er, S. Wu, J. Lu, and H. L. Toh, "Face recognition with radial basis function (RBF) neural networks", *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 697–710, 2002.
- [66] V. N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, 1998.
- [67] Y. M. Li, S. G. Gong, and H. Liddell, "Support vector regression and classification based multi-view face detection and recognition", in *International Conference on Face and Gesture Recognition*, France, 2000, pp. 300–305.
- [68] G. Guo, S. Z. Li, and K. L. Chan, "Face recognition by support vector machines", in *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 196–201.
- [69] B. Heisele, P. Ho, and T. Poggio, "Face recognition with support vector machines: Global versus component-based approach", in *Proceedings of the International Conference on Computer Vision*, 2001, pp. 688–694.
- [70] G. L. Giles and T. Maxwell, "Learning, invariances, and generalization in high-order neural networks", *Applied Optics*, vol. 26, pp. 4972–4978, 1987.
- [71] G. L. Giles, R. D. Griffin, and T. Maxwell, "Encoding geometric invariances in higher-order neural networks", in *Neural Information Processing Systems, American Institute of Physics Conference Proceedings*, 1988, pp. 301–309.

- [72] M. B. Reid, L. Spirkovska, and E. Ochoa, "Rapid training of higher-order neural networks for invariant pattern recognition", in *Proceedings of Joint International Conference on Neural Networks*, 1989, pp. 689–692.
- [73] L. Spirkovska, M. B. Reid, and E. Ochoa, "Simultaneous position, scale, and rotation invariant pattern classification using third-order neural networks", *International Journal of Neural Networks*, vol. 1, pp. 154–159, 1989.
- [74] S. J. Perantonis and P. J. G. Lisboa, "Translation, rotation, and scale invariant pattern recognition by high-order neural networks and moment classifier", *IEEE Transactions on Neural Networks*, vol. 3, no. 2, pp. 241–251, 1992.
- [75] L. Spirkovska and M. B. Reid, "Coarse-coded higher-order neural networks for PSRI object recognition", *IEEE Transactions on Neural Networks*, vol. 4, no. 2, pp. 276–283, 1993.
- [76] L. Spirkovska and M. B. Reid, "Higher-order neural networks applied to 2d and 3d object recognition", *Machine Learning*, vol. 15, pp. 169–199, 1994.
- [77] O. A. Uwechue and A. S. Pandya, *Human Face Recognition Using Third-Order Synthetic Neural Networks*, Boston : Kluwer Academic Publishers, 1997.
- [78] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification", in *Proceedings of IEEE Workshop on Application of Computer Vision*, Sarasota, FL.
- [79] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 630–660, 2001.

- [80] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database"; *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, 2003.
- [81] A. M. Martinez and R. Benavente, "The AR face database", *CVC Technical Report #24*, 1998.
- [82] D. B. Graham and N. M. Allinson, "Characterizing virtual eigensignatures for general purpose face recognition", *Face Recognition: From Theory to Applications, NATO ASI Series F, Computer and Systems Sciences*, pp. 446–456, 1998.
- [83] K. Messer, J. Matas, J Kittler, J Luettin, and G maitre, "XM2VTSDB: The extended M2VTS database", in *IAPR International Conference on Audio- and Video-based Biometric Person Authentication*, March.
- [84] K. Fukunaga, *Introduction to Statistical Pattern Recognition, Second Edition*, Academic Press, 1990.
- [85] Z. Pan, R. Adams, and H. Bolouri, "Image redundancy reduction for neural network classification using discrete cosine transforms", in *Proceedings of The IEEE-INNS-ENNS International Joint Conference on Neural Networks*, Como, Italy, 2000, vol. 3, pp. 149–154.
- [86] Z. M. Hafeed and M. D. Levine, "Face recognition using the discrete cosine transform", *International Journal of Computer Vision*, vol. 43, no. 3, pp. 167–188, 2001.
- [87] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*, Academic Press, 1990.
- [88] J. Zhu, M. I. Vai, and P. U. Mak, "Face recognition using 2D DCT with PCA", in *Proceedings of The 4th Chinese Conference on Biometric Recognition (Sinobiometrics'03)*, Beijing, P. R. China, 2003.

- [89] W. Pennebaker and J. Mitchell, *JPEG Still Image Data Compression Standard*, Van Nostrand Reinhold, New York, 1993.
- [90] R. Beveridge, D. Bolme, M. Teixeira, and B. Draper, *The CSU Face Identification Evaluation System User's Guide: Version 5.0*, Colorado State University, 2003.
- [91] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, 1992.
- [92] F. Yang and M. Paindavoine, "Implementation of an rbf neural network on embedded systems: Real-time face tracking and identity verification", *IEEE Transactions on Neural Networks*, vol. 14, no. 5, pp. 1162–1175, 2003.
- [93] J. Park and J. Wsandberg, "Universal approximation using radial basis functions network", *Neural Computation*, vol. 3, pp. 246–257, 1991.
- [94] S. Lee and R. M. Kil, "A gaussian potential function network with hierarchically self-organizing learning", *Neural Networks*, vol. 4, pp. 207–224, 1991.
- [95] J. Moody and C. J. Darken, "Fast learning in network of locally-tuned processing units", *Neural Computation*, vol. 1, pp. 281–294, 1989.
- [96] W. Pedrycz, "Conditional fuzzy clustering in the design of radial basis function neural networks", *IEEE Transactions on Neural Networks*, vol. 9, pp. 601–612, 1998.
- [97] S. Wu and M. J. Er, "Dynamic fuzzy neural networks: A novel approach to function approximation", *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 30, no. 2, pp. 358–364, 2000.
- [98] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 974–989, 1999.

- [99] A. G. Bors and I. Pitas, "Median radial basis function neural network", *IEEE Transactions on Neural Networks*, vol. 7, pp. 1351–1364, 1996.
- [100] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, 1992.
- [101] A. S. Tolba and A. N. Abu-Rezq, "Combined classifiers for invariant face recognition", in *Proceedings of International Conference on Information Intelligence and Systems*, Bethesda, Maryland, USA, 1999.
- [102] T. Tan and H. Yan, "Object recognition based on fractal neighbor distance", *Signal Processing*, vol. 81, pp. 2105–2129, 2001.
- [103] Y.-S. Ryu and S.-Y. Oh, "Simple hybrid classifier for face recognition with adaptively generated virtual data", *Pattern Recognition Letters*, vol. 23, pp. 833–841, 2002.
- [104] S. Shan, W. Gao, B. Cao, and D. Zhao, "Illumination normalization for robust face recognition against varying lighting conditions", in *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, Nice, France, 2003.
- [105] M. Savvides and V. Kumar, "Illumination normalization using logarithm transforms for face authentication", in *IAPR International Conference on Audio- and Video-based Biometric Person Authentication*, Guildford, UK, 2003.
- [106] X. Xie and K.-M. L, "Face recognition under varying illumination based on a 2D face shape model", *Pattern Recognition*, vol. 38, no. 2, pp. 355–368, 2005.
- [107] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible illumination conditions", *International Journal of Computer Vision*, vol. 28, no. 3, pp. 245–260, 1998.

- [108] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 721–732, 1997.
- [109] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, 2003.
- [110] A. Shashua and T. Riklin-Raviv, "The quotient image: Class-based re-rendering and recognition with varying illuminations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 129–139, 2001.
- [111] W. Zhao and R. Chellappa, "Illumination-insensitive face recognition using symmetric shape-from-shading", in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2000, pp. 286–293.
- [112] L. Zhang and D. Samaras, "Face recognition under variable lighting using harmonic image exemplars", in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003, vol. 1, pp. 19–25.
- [113] J. Zhao, D. Wang Y. Su, and S. Luo, "Illumination ratio image: Synthesizing and recognition with varying illuminations", *Pattern Recognition Letters*, vol. 24, pp. 2703–2710, 2003.
- [114] S.M. Pizer and E.P. Amburn, "Adaptive histogram equalization and its variations", *Computer Vision, Graphics and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.
- [115] B. K. P. Horn, *Robot Vision*, MIT Press, Cambridge, Massachusetts, 1986.
- [116] E. H. Land and J. J. McCann, "Lightness and retinex theory", *Journal of the Optical Society of America*, vol. 61, pp. 1–11, 1971.

- [117] S. K. Nayar and R. M. Bolle, "Reflectance based object recognition", *International Journal of Computer Vision*, vol. 17, no. 3, pp. 219–240, 1996.
- [118] C.-H. Lin and J.-L. Wu, "Automatic facial feature extraction by genetic algorithms", *IEEE Transactions on Image Processing*, vol. 8, no. 6, pp. 834–845, 1999.
- [119] T. Kawaguchi, D. Hidaka, and M. Rizon, "Detection of eyes from human faces by hough transform and separability filter", in *Proceedings of International Conference on Image Processing*, 2000, vol. 1, pp. 49–52.
- [120] T. Kawaguchi, D. Hidaka, and M. Rizon, "Robust extraction of eyes from face", in *Proceedings of International Conference Pattern Recognition*, 2000, vol. 1, pp. 1109–1114.
- [121] J. Wood, "Invariant pattern recognition: A review", *Pattern Recognition*, vol. 29, no. 1, pp. 1–17, 1996.
- [122] Y. Sheng and H. Arsenault, "Experiments on pattern recognition using invariant fourier-mellin descriptors", *Journal of the Optical Society of America A*, vol. 3, no. 6, pp. 771–776, 1986.
- [123] A. Delopoulos, A. Tirakis, and S. Kollias, "Invariant image classification using triple-correlation-based neural networks", *IEEE Transactions on Neural Networks*, vol. 5, no. 3, pp. 392–408, 1994.
- [124] M. K. Hu, "Visual pattern recognition by moment invariants", *IEEE Transactions on Information Theory*, vol. 8, pp. 179–187, 1962.
- [125] A. Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 489–497, 1990.
- [126] J. H. Lai, P. C. Yuen, and G. C. Feng, "Face recognition using holistic fourier invariant features", *Pattern Recognition*, vol. 34, no. 1, pp. 95–109, 2001.

- [127] J. Haddadnia and M. Ahmadi, "N-feature neural network human face recognition", *Image and Vision Computing*, vol. 22, no. 12, pp. 1071–1082, 2004.
- [128] T. Kanaoka, R. Chellappa, M. Yoshitaka, and S. Tomita, "A higher-order neural network for distortion invariant pattern recognition", *Pattern Recognition Letters*, vol. 13, pp. 837–841, 1992.
- [129] M. Takano, T. Kanaoka, J. Skrzypek, and S. Tomita, "A note on a higher-order neural network for distortion invariant pattern recognition", *Pattern Recognition Letters*, vol. 15, pp. 631–635, 1994.
- [130] J. M. Mendel, "Tutorial in higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications", *Proceedings of the IEEE*, vol. 79, pp. 278–305, 1991.