IMAGE-BASED FACE RECOGNITION UNDER VARYING POSE AND ILLUMINATION CONDITIONS

by

Shan Du

A THESIS SUBMITTED IN PARTIAL FULLFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

November 2008

ABSTRACT

Image-based face recognition has attained wide applications during the past decades in commerce and law enforcement areas, such as mug shot database matching, identity authentication, and access control. Existing face recognition techniques (e.g., Eigenface, Fisherface, and Elastic Bunch Graph Matching, etc.), however, do not perform well when the following case inevitably exists. The case is that, due to some variations in imaging conditions, e.g., pose and illumination changes, face images of the same person often have different appearances. These variations make face recognition techniques much challenging. With this concern in mind, the objective of my research is to develop robust face recognition techniques against variations.

This thesis addresses two main variation problems in face recognition, i.e., pose and illumination variations. To improve the performance of face recognition systems, the following methods are proposed: (1) a face feature extraction and representation method using non-uniformly selected Gabor convolution features, (2) an illumination normalization method using adaptive region-based image enhancement for face recognition under variable illumination conditions, (3) an eye detection method in gray-scale face images under various illumination conditions, and (4) a virtual pose generation method for pose-invariant face recognition. The details of these proposed methods are explained in this thesis. In addition, we conduct a comprehensive survey of the existing face recognition methods. Future research directions are pointed out.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D	2-Dimension
3D	3-Dimension
AAM	Active Appearance Model
AdaBoost	Adaptive Boosting
AGFC	Adaboost Gabor Fisher Classifier
ARHE	Adaptive Region-based Histogram Equalization
ASM	Active Shape Model
BHE	Block-based Histogram Equalization
CMU-PIE	Carnegie Mellon University Pose, Illumination and Expression Database
DCM	Dynamic Committee Machine
DLA	Dynamic Link Architecture
DWT	Discrete Wavelet Transform
EBGM	Elastic Bunch Graph Matching
EdgeE	Edge Enhancement
EER	Equal Error Rate
EGM	Elastic Graph Matching
ELF	Eigen Light-Field
EP	Evolution Pursuit
FERET	Facial Recognition Technology Database
FLD	Fisher Linear Discriminant
FRVT	Face Recognition Vendor Test
GA	Genetic Algorithm
GBR	Generalized Bas Relief
GFC	Gabor Fisher Classifier
GIC	Gamma Intensity Correction
GLR	Global Linear Regression
GPF	Generalized Projection Function
GWN	Gabor Wavelet Network
HE	Histogram Equalization
HPF	Hybrid Projection Function
HRI	Human Robot Interaction
ICA	Independent Component Analysis
ICP	Iterative Closest Point
IDWT	Inverse Discrete Wavelet Transform
IPF	Integral Projection Function
IR	Infrared
KLT	Karhunen-Loeve Transform
LDA	linear Discriminant Analysis
LFA	Local Feature Analysis
LLR	Local Linear Regression
LOC	Linear Object Class
MNFL	Modular Nearest Feature Line

MSE	Mean Square Error
NUGFs	Non-Uniform Gabor Features
ORL Database	Olivetti Research Laboratory Database
PCA	Principal Component Analysis
QIR	Quotient Illumination Relighting
RHE	Region-based Histogram Equalization
SSFS	Symmetric Shape-From-Shading
TLNN	Two-Layer Nearest Neighbor
VPF	Variance Projection Function
Yale B	Yale Face Database B

ACKNOWLEDGEMENTS

This dissertation is the fruit of five years of research. I would like to express my sincere appreciation to all those who have supported me for completing my study.

First, I would like to offer my enduring gratitude to my research supervisor, Prof. Rabab K. Ward, for providing me the opportunity to work in her research group. Her constant support, encouragement, patience and invaluable guidance on my research and this dissertation are highly appreciated. Without her supervision, this dissertation would not have been possible. Next, I extend my thanks to the researchers in the Image Processing Lab of UBC for their support and help.

I gratefully acknowledge the Department of Electrical and Computer Engineering, the University of British Columbia, and Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support. I would like to thank NSERC for providing me the Canada Graduate Scholarship (CGS D) that covered my living expenses.

Special thanks are owed to my parents, who have supported me throughout my years of education. I want to express my gratitude to them. Their support, love, and encouragement made this dissertation possible. I cannot stress enough how thankful I am to my husband Qixiang for his love, understanding, encouragement and endless support. None of this would be possible without his help.

I would also like to express my gratitude to each of the committee members for his/her precious time and advice.

To my parents,
my husband and our daughter

CO-AUTHORSHIP STATEMENT

Manuscript 1: Shan Du and Rabab Ward, "Improved Face Representation by Non-uniform Multi-level Selection of Gabor Convolution Features," *IEEE Transactions on Systems, Man and Cybernetic, Part B*, accepted for publication.

Shan Du conducted the research, proposed the method, implemented the experiments, wrote the manuscript and acted as the corresponding author.

Rabab Ward supervised the research and the writing of the manuscript.

Manuscript 2: Shan Du and Rabab Ward, "Adaptive Region-based Image Enhancement Method for Robust Face Recognition under Variable Illumination Conditions," *IEEE Transactions on Circuits and Systems for Video Technology*, submitted.

Shan Du conducted the research, proposed the method, implemented the experiments, wrote the manuscript and acted as the corresponding author.

Rabab Ward supervised the research and the writing of the manuscript.

Manuscript 3: Shan Du and Rabab Ward, "Face Recognition under Pose Variations," *Journal of the Franklin Institute*, vol. 343, no. 6, pp. 596-613, 2006.

Shan Du reviewed the papers, created the tables, interpreted the results, wrote the manuscript and acted as the corresponding author.

Rabab Ward Rabab Ward supervised the research and the writing of the manuscript.

Manuscript 4: Shan Du and Rabab Ward, "Eye Detection in Gray-scale Face Images under Various Illumination Conditions," *IEEE Transactions on Information Forensics and Security*, submitted.

Shan Du conducted the research, proposed the method, implemented the experiments, wrote the manuscript and acted as the corresponding author.

Rabab Ward supervised the research and the writing of the manuscript.

Manuscript 5: Shan Du and Rabab Ward, "Facial-Component-Wise Pose Normalization for Pose-Invariant Face Recognition," *IEEE Transactions on Multimedia*, submitted.

Shan Du conducted the research, proposed the method, implemented the experiments, wrote the manuscript and acted as the corresponding author.

Rabab Ward supervised the research and the writing of the manuscript.

CHAPTER 1 THESIS OVERVIEW

1.1 Introduction

In recent years, the need for accurate and automatic human recognition techniques has seen much growth. Face recognition is a form of biometrics that can assist in the human recognition process. Biometrics uses distinguishable forms of the human anatomy (physical characteristics) or traits (behavioral characteristics) to determine or verify the identity of an individual. Physical characteristics include facial patterns, fingerprints, eye retinas and irises, DNA, and hand geometry. Behavioral characteristics, on the other hand, include signature, gait, voice and typing patterns.

Among the available biometrics methods, such as, face recognition, iris recognition, fingerprint matching, and DNA matching, face recognition is much more desirable and has the most applications. Therefore, the focus of my research and the thesis is on face recognition technology.

Face recognition has a wide range of applications, from identity authentication, mug shot matching, access control, and face-based video indexing/browsing, to human-computer interaction.

Face recognition technology is the least intrusive and fastest biometric technology. For example in surveillance systems, instead of requiring people to place their hands on a reader (fingerprinting) or precisely position their eyes in front of a scanner (iris recognition), face recognition systems unobtrusively take pictures of people's faces as they enter a defined area. There is no intrusion or capture delay, and in most cases, the subjects are entirely unaware of the process. People do not necessarily feel "under surveillance" or their privacy being invaded.

Building an automatic face recognition system has been an active research topic in computer vision and pattern recognition for few decades. A general statement of the face recognition problem is simply formulated as follows: given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces. The solution to the problem involves three steps in general: (1) segmentation of faces (face detection) from cluttered scenes, (2) feature extraction from the face regions, and (3) face classification or verification [1]-[3].

Even though humans can detect and identify faces in a scene with little effort, building an automated system that accomplishes such an objective is not that simple. The challenges are even more profound when one considers the large variations in visual stimulus. These variations become the concern of my research.

Variations associated with face images are attributed to the following factors:

- Pose - The images of a face vary due to the position of the face relative to the camera (e.g., frontal, 45 degree, profile, and upside down).
- Illumination and imaging conditions - When an image is formed, factors such as lighting (spectra, source distribution and intensity) and camera characteristics (sensor response, lenses) affect the appearance of a face.
- Facial expression - The appearance of faces is directly affected by a person's facial expression.
- Scale - Face images may have different sizes.
- Occlusion - Some facial features such as beards, mustaches, and glasses may be present in some pictures of the same person. In these pictures, some facial characteristics get occluded. Faces may also be partially occluded by other objects.

Based on the above observation, my research mainly focuses on proposing face recognition techniques that are robust against the two most significant variations involved in face images, pose and illumination variations.

1.2 Literature Review

There are many closely related problems of face recognition. Among them, face localization aims to determine the image position of a single face. The goal of facial feature detection is to detect the presence and location of features, such as eyes, nose, mouth, etc. The purpose of face authentication is to verify the identity claim of an individual in an input image, while face tracking continuously estimates the location and possibly the orientation of a face in an image sequence in real-time. Facial expression recognition is concerned with identifying the affective states (e.g., happy, sad, surprised, disgusted) of humans. This section gives an overview of the related work.

The earliest work on face recognition can be traced back at least to the 1950s in the psychology literature [4] and to the 1960s in engineering [5]. But research on automatic machine recognition of faces started in the 1970s [6][7]. Earlier approaches treated face recognition as a 2D pattern recognition problem. As a result, during the early and mid-1970s, pattern classification techniques were used to measure attributes (e.g., the distances between important points) in faces or face profiles. During the 1980s, work on face recognition remained largely dormant. Since the early 1990s, research interest in face recognition has grown significantly due to the interest in commercial opportunities, the availability of real-time hardware, and the increasing importance of surveillance-related applications.

Existing face recognition approaches can be categorized into the following categories:

(1) Holistic matching approaches. These approaches use the whole face region as the

raw input to a recognition system.

(2) Feature-based matching approaches. Features studied are local features such as lines or fiducial points, or facial features such as eyes and mouth. In these approaches, features are first extracted and their locations and local statistics (geometric and/or appearance) are fed into a classifier. The detection of faces and their features prior to performing verification or recognition makes these approaches robust to positional variations of the faces in the input image.

1.2.1 Holistic Approaches

One of the most successful holistic approaches is the principal component analysis (PCA) or the Karhunen-Loeve transform (KLT). Kirby and Sirovich demonstrated that the images of faces could be linearly encoded using a modest number of basis images [8]. This is based on the KL transform, which also goes by other names, e.g., principal component analysis [9], and the Hotelling transform [10]. KLT completely decorrelates a signal in the transform domain, minimizes the mean square error (MSE) in data compression, contains the most energy in the fewest number of transform coefficients, and minimizes the total representation entropy of the input sequence [11]. All of these properties are extremely useful in pattern recognition applications. The computation of the KLT essentially involves the determination of the eigenvectors of the covariance matrix of a set of training images. Given a collection of n by m pixel training images, each image is first represented as a vector of size $m \times n$. The basis vectors spanning an optimal subspace are then determined such that the mean square error between the projection of the training images into this subspace and the original images is minimized. The optimal basis vectors are called eigenpictures since these are the

eigenvectors of the covariance matrix of the vectors of the face images in the training set. These eigenpictures later became known as Eigenfaces [12].

An advantage of using such representations is their reduced sensitivity to noise. Some of this noise may be due to small occlusions, as long as the topological structure does not change. However, KLT does not achieve adequate robustness against variations in face orientation, position, and illumination. That is why it is usually accompanied by further processing to improve its performance.

In mug shot applications, a frontal and a side view of a person are usually available. To handle images from multiple views, two approaches can be taken [13]. The first approach pools all the images and constructs a set of eigenfaces that represent all the images from all the views. The other approach uses separate eigenfaces for different views, so that the collection of images taken from each view has its own eigenspace. The second approach, known as the view-based eigenspaces, performs better.

Using a probabilistic measure of similarity, instead of the simple Euclidean distance used with eigenfaces, the standard eigenface approach was extended to a Bayesian approach [14]. Practically, the major drawback of the Bayesian approach is the need to estimate the probability distributions in a high-dimensional space from very limited numbers of training samples per class.

Another successful holistic approach to face recognition is based on the linear/Fisher discriminant analysis (LDA/FLD) [15]. While eigenfaces suffer from lighting and facial expression variations, Fisherface is insensitive to large variations in the lighting direction and facial expression. In this approach, the Fisher's linear discriminant is used to obtain the most discriminating features in faces, rather than the most expressive ones given by KLT.

Fisherface maximizes the ratio of the between-class scatter to the within-class scatter. Fisherface is one of the most successful face recognition methods, but unfortunately, it requires several training images for each face. So it cannot be applied to face recognition applications where only one sample image per person is available for training. To tackle this problem, the Fisherface approach was extended by deriving multiple images of a face from one single image. Fisherface was then trained on these derived images [16].

The evolution pursuit (EP) approach implements strategies based on the characteristic of genetic algorithms (GAs) for searching the space of possible solutions to determine the optimal basis [17]. EP starts by projecting the original data into a lower-dimensional whitened PCA space. Directed random rotations of the basis vectors in this space are then searched by GAs where evolution is driven by a fitness function defined in terms of performance accuracy and class separation.

Based on the argument that for tasks such as face recognition much of the important information is contained in high-order statistics, a method that uses independent component analysis (ICA) to extract features for face recognition has been proposed [18]. ICA is a generalization of principal component analysis, which decorrelates the high-order moments of the input in addition to the second-order moments. Two architectures have been proposed for face recognition. In both architectures, PCA is first used to reduce the dimensionality of the image size.

1.2.2 Feature-based Approaches

Lades *et al*. used an artificial neural network, which employs the so-called dynamic link architecture (DLA), to achieve distortion-invariant recognition [19]. Local descriptors of the input images are obtained using Gabor-based wavelets. By conveying frequency, position,

and orientation information, this approach performs well on relatively large databases. For a practical implementation of the dynamic link matching, elastic graph matching (EGM) has been proposed [20][21]. This is a neural network with dynamically evolving links between a reference model and an input model image. To characterize a face, the EGM method utilizes an attributed relational graph, with facial landmarks (fiducial point) as the graph nodes, the Gabor transform around each fiducial point as the node attributes or jets and the distances between nodes as edge attributes. To compute the jet values, the fiducial points have to be located first. The is done through an elastic graph matching process, where the nodes of a model graph are tentatively overlaid on the test image, and the jets are extracted from the local image area around each node. Then the similarity of the model graph and the test image graph is optimized by dynamically varying the node positions in the image until the best matching location is found. Each time when a node location is changed, the jet value of the node has to be recomputed through the Gabor transform. Hence, the elastic matching process is very time consuming. This limits the EGM method in many practical applications.

Elastic bunch graph matching (EBGM) recognizes a human face from a large database containing one image per person [22]. This method differs from EGM in three aspects. First, the phase of the complex Gabor wavelet coefficients is used to achieve a more accurate location of the nodes and to lessen the ambiguity in patterns that are similar in their coefficient magnitudes. Secondly, object-adapted graphs are employed, so that nodes refer to specific facial landmarks, called fiducial points. The correct correspondences between two faces can then be found across large viewpoint changes. Thirdly, a new data structure, called the bunch graph, is introduced; this graph serves as a generalized representation of faces by

combining jets of a small set of individual faces. The success of EBGM is due to its resemblance to the human visual system.

1.2.3 Hybrid Approaches

There are many approaches that use both holistic and local features. We call them hybrid approaches. For example, the modular eigenface approach uses both global eigenfaces and local eigenfeatures [13].

A method that is based on a flexible appearance model for automatic face recognition has been presented in [23]. To identify a face, both shape and gray-level information are modeled and used. The shape model is an active shape model (ASM), which is a statistical model of the shapes of objects [24]. It can iteratively deform to fit to an example of the shape in a new image. The statistical shape model is trained on sample images using PCA, where the variables are the coordinates of the shape model points. For the purpose of classification, the shape variations due to inter-class variation are separated from those due to within-class variations using discriminant analysis. Based on the average shape of the shape model, a global shape-free gray-level model can be constructed, again using PCA. To further enhance the robustness of the system against changes in local appearance such as occlusions, local gray-level models are also built on the shape model points. Simple local profiles perpendicular to the shape boundary are used. Finally, for an input image, all three types of information, including extracted shape parameters, shape-free image parameters, and local profiles, are used to compute a Mahalanobis distance for classification.

In [25], a method based on component-based detection/recognition [26] and 3D morphable models [27] was presented. The basic idea of component-based methods is to decompose a face into a set of facial components such as mouth and eyes that are

interconnected by a flexible geometrical model. The motivation for using facial components is that changes in head pose mainly lead to changes in the positions of facial components, and these changes could be accounted for by the flexibility of the geometric model. A major drawback of the system, however, is its need of a large number of training images taken from different viewpoints and under different lighting conditions. To overcome this problem, the 3D morphable face model was applied to generate arbitrary synthetic images under varying pose and illumination. Only three face images of a person are needed to compute the 3D face model. Once the 3D model is constructed, synthetic images are generated for training both the detector and the classifier.

Hybrid approaches that use both holistic and local features seem to be promising since they resemble the human perceptual system. While the holistic approach provides a quick recognition method, the discriminant information that it provides may not be rich enough to handle very large databases. This insufficiency can be compensated for by local feature methods.

The above gives a brief overview of the existing work on face recognition. In Chapter 2 to 6, more detailed reviews of the existing work related to their respective topics will be presented.

Two major problems regarding face recognition have not been well solved by the exiting face recognition methods listed above, i.e., face recognition under varying pose and illumination conditions. Thus, the two major obstacles lead to the objective of my research.

1.3 Research Objective

The objective of my research is to investigate and propose appropriate face recognition techniques that are robust against the two most significant variations involved in face images, pose and illumination variations.

1.4 Research Approaches and Major Contributions

In order to achieve the objective, we tackle the problems from two aspects: (1) to propose a new face feature extraction and representation technique, and (2) to propose new problem-specific pre-processing techniques. Thus, this thesis describes new methodologies to make automatic face recognition systems independent from the acquisition conditions of the images (mainly geometrical transforms and illumination), as well as a new technique for automatic extraction of highly discriminating characteristics from human face images.

The performance of a face recognition system highly depends on the representation of the face patterns (i.e., feature extraction). Generally speaking, a good representation should have characteristics such as: (1) small within-class variations, (2) large between-class variations, and (3) low-dimensional space (i.e., short vector length) in order to avoid the high computational cost in the classifier. Furthermore, its extraction should not depend much on manual operations. Intuitively, one should derive a face representation from the 3D face shape and skin reflectance if we could recover the above intrinsic information from a given 2D face image. Unfortunately, this is an ill-posed problem in computer vision. Therefore, most current well-known face recognition methods derive a face representation directly from the 2D face image matrix.

Another popular strategy for representing face patterns is to exploit some mathematical transformations of the 2D image. Typical transformations include the Fourier transform and

various wavelet transforms. Among them, Gabor wavelets have been widely accepted by researchers in face recognition community, mostly because its kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells and exhibit desirable characteristics of spatial locality and orientation selectivity. Previous work on Gabor features has also demonstrated excellent performance. Typical methods include the dynamic link architecture (DLA) [19], elastic graph matching (EGM) [20], Gabor wavelet network (GWN) [28], and Gabor-Fisher classifier (GFC) [29].

As an outcome of the first part of my research, a new feature extraction and representation technique has been proposed. In this thesis, *we present a new face representation method that employs non-uniform multi-level selection of Gabor features*. The proposed face representation method has the advantages of low complexity, low-dimensionality and high discriminance. The proposed method works well for cases when multiple sample images are available for each person for training as well as when only one sample image is available for each person.

When the face images are highly varied due to severe pose and illumination changes, the sole dependency on feature extraction is not enough. We strongly believe that adding more delicately designed pre-processing methods before feature extraction can substantially increase face recognition performance in cases of heavy variations, which is the focus of the second part of my research. We propose problem-specific pre-processing methods to deal with different variations.

Variable illumination conditions in face images, especially the side lighting effect, form a main obstacle in face recognition systems. In this thesis, *for face recognition under variable illumination conditions, an illumination normalization method using adaptive*

region-based image enhancement is proposed. The proposed method does not require any 3D modeling and model fitting steps and can be easily implemented. It can be applied directly to any single image without using prior lighting assumptions, nor any prior information on 3D face geometry.

Localization of eyes is a necessary step for many face recognition systems. Before two face images can be compared, they should be aligned in orientation and normalized in scale. Since both the locations of the two eyes and the interocular distance are relatively constant for most people, the eyes are often used for face image normalization. In this thesis, *a new method of automatic localization of eyes in face images that is invariant to pose and illumination changes is presented*.

The pose variation involved in face images significantly degrades the performance of face recognition systems. After a comprehensive survey of the existing methods dealing with the pose variations including both 2D and 3D methods, in this thesis, *a novel facial-component-wise virtual pose generation method for facilitating pose-invariant face recognition is proposed*. With this efficient facial-component-based pose normalization method, both visual quality and recognition rate are shown to increase.

In summary, in order to achieve the research objective, we have proposed: (1) a new face representation method, (2) a new illumination normalization method, (3) a new facial feature localization method, and (4) a new face pose generation method. With these new methods, we are able to substantially improve the performance of an automatic face recognition system in terms of recognition rate, processing speed and resource consumption. The improvements have been proven by numerous experimental results on the most popular face databases such as ORL, Yale, FERET, Yale B and CMU-PIE.

The proposed methods can act to improve performance independently or to collaborate as a whole system to deal with more complex cases. Development of a complete face recognition software tool integrating all these methods is under plan and will be a promising and worthwhile work to do after my PhD research.

1.5 Organization of the Thesis

The organization of this thesis is explained as follows.

In Chapter 2, we propose a new face representation method that employs a non-uniform multi-level selection of Gabor features. The proposed method is based on the local statistics of the Gabor features and is implemented using a coarse-to-fine hierarchical strategy. Gabor features that correspond to important face regions are automatically selected and sampled finer than other features. The non-uniformly extracted Gabor features are then classified using principal component analysis and/or linear discriminant analysis for the purpose of face recognition. To verify the effectiveness of the proposed method, experiments have been conducted on benchmark face image databases where the images vary in illumination, expression, pose, and scale. Compared with the methods that use the original gray-scale image with 4096-dimension data and uniform sampling with 2560-dimension data, the proposed method results in a significantly higher recognition rate, with a substantial lower dimension of around 700. The experimental results also show that the proposed method works well not only when multiple sample images are available for training but also when only one sample image is available for each person. The proposed face representation method has the advantages of low complexity, low-dimensionality and high discriminance.

In Chapter 3, we address the illumination variation problem. Variable illumination conditions, especially the side lighting effect in face images form a main obstacle in face

recognition systems. To deal with this problem, this chapter presents a novel adaptive region-based image preprocessing scheme that enhances face images and facilitates the illumination invariant face recognition task. The proposed method first segments an image into different regions according to its different local illumination conditions, then both the contrast and the edges are enhanced on a region by region basis so as to alleviate the side lighting effect. Different from all existing contrast enhancement methods, we apply the proposed adaptive region-based histogram equalization (ARHE) on the low-frequency coefficients to minimize the illumination variations under different lighting conditions. Besides contrast enhancement, by observing that under poor illuminations the high-frequency features become more important in recognition, we propose to enlarge the high-frequency coefficients to make face images more distinguishable. This procedure is called edge enhancement (EdgeE). The edge enhancement is also region-based. Compared with existing image preprocessing methods, our method is shown to be more suitable for dealing with uneven illuminations in face images. Experimental results show that the proposed method significantly improves the recognition performance of face images with illumination variations. The proposed method does not require any 3D modeling and model fitting steps and can be easily implemented. Moreover, it can be applied directly to any single image without using any lighting assumption or any prior information on 3D face geometry.

In Chapter 4, we propose a novel automated method that localizes the eyes in gray-scale face images and is robust to illumination variations. The method does not require prior knowledge about face orientation and illumination strength. Other advantages are that manual initialization or training process is not needed. This method consists of four steps. Based on an edge map obtained via a multi-resolution wavelet transform, the method first

segments an image into different heterogeneously illuminated regions. The illumination of every region is then adjusted separately so that the features' details are more pronounced. To locate the different facial features, for every region, a Gabor-based image is constructed from the illumination adjusted image. The eyes sub-regions are then identified using the edge map of the illumination adjusted image. This method has been successfully applied to the images of the Yale B face database that have different illuminations and different poses.

Chapter 5 and Chapter 6 address one of the main obstacles in the face recognition task, the variations in face pose. In Chapter 5, we present a comprehensive review of the typical algorithms that aim to overcome this problem. These algorithms are categorized and briefly described. Future research challenges in pose-invariant face recognition are also identified. Then in Chapter 6, a novel facial-component-wise pose normalization method for facilitating the pose-invariant face recognition is proposed. The main idea is to normalize a non-frontal face image to a virtual frontal image component by component. In this method, we first partition the whole non-frontal face image into different facial components and then the virtual frontal view for each component is estimated separately. The final virtual frontal image is generated by integrating the virtual frontal components. The proposed method relies only on 2D images, therefore complex 3D modeling is not needed. Experimental results using the CMU-PIE database demonstrate the advantages of the proposed method over the local linear regression (LLR) method and the eigen light-field (ELF) method.

In Chapter 7, we summarize the contributions of this thesis to the face recognition field. We also present some of the potential research subjects that can follow this research.

1.6 References

[1] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, 2003.

[2] R. Chellappa, C. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proc. of the IEEE*, vol. 83, no. 5, pp. 705-740, 1995.

[3] M. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34-58, 2002.

[4] I. Bruner and R. Tagiuri, "The perception of people," *Handbook of Social Psychology*, vol. 2, G. Lindzey, Ed., Addison-Wesley, Reading, MA, pp. 634-654, 1954.

[5] W. Bledsoe, "The model method in facial recognition," *Tech. Rep. PRI-15, Panoramic research Inc.*, Palo Alto, CA, 1964.

[6] M. Kelly, "Visual identification of people by computer," *Tech. Rep. AI-130, Stanford AI Project*, Stanford, CA, 1970.

[7] T. Kanade*, Computer Recognition of Human Face*, Birkhauser, Basel, Switzerland, and Stuttgart, Germany, 1973.

[8] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103-108, 1990.

[9] L. Jolliffe, *Principal Component Analysis*, New York: Springer-Verlag, 1986.

[10] R. Gonzalez and P. Wintz, *Digital Image Processing*, Reading: Addison Wesley, 1987.

[11] A. Rosenfeld and A. Kak, *Digital Picture Processing*, Academic: New York, NY, 1976.

[12] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, pp. 72-86, 1991.

[13] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenfaces for face recognition," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 84-91, 1994.

[14] B. Moghaddam, C. Nastar, and A. Pentland, "A Bayesian similarity measure for direct image matching," *Proc. of International Conference on Pattern Recognition*, vol. 2, pp. 350-358, 1996.

[15] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.

[16] S. Shan, B. Cao, W. Gao, and D. Zhao, "Extended Fisherface for face recognition from a single example image per person," *Proc. of the IEEE*, vol. 2, pp. 81-84, 2002.

[17] C. Liu and H. Wechsler, "Evolutionary pursuit and its application to face recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 570-582, 2000.

[18] M. Bartlett, H. Lades, and T. Sejnowski, "Independent component representation for face recognition," *Proc. of SPIE Symposium on Electronic Imaging: Science and Technology*, pp. 528-539, 1998.

[19] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. on Computers*, vol. 42, pp. 300-311, 1993.

[20] B. Duc, S. Fischer, and J. Bigun, "Face authentication with gabor information on deformable graphs," *IEEE Trans. on Image Processing*, vol. 8, no. 4, pp. 504-516, 1999.

[21] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal face authentication using morphological elastic graph matching," *IEEE Trans. on Image Processing*, vol. 9, no. 4, pp. 555-560, 2000.

[22] L. Wiskott, J. Fellous, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 775-779, 1997.

[23] A. Lanitis, C. Taylor, and T. Cootes, "Automatic face identification system using flexible appearance models," *Image and Vision Computing,* vol. 13, pp. 393-401, 1995.

[24] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models - their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38-59, 1995.

[25] J. Huang, B. Heisele, and V. Blanz, "Component-based face recognition with 3D morphable models," *Proc. of International Conference on Audio- and Video-based Person Authentication*, 2003.

[26] B. Heisele, T. Serre, M. Pontil, and T. Poggio, "Component-based face detection," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[27] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," *Proc. of SIGGRAPH'99*, pp. 187-194, 1999.

[28] V. Kruger and G. Sommer, "Gabor wavelet networks for object representation," Tech. *Rep. 2002*, Institute of Computer Science, University of Kiel, 2000.

[29] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. on Image Processing*, vol. 11, no. 4, pp. 467-476, 2002.

CHAPTER 2 IMPROVED FACE REPRESENTATION BY NON-UNIFORM MULTI-LEVEL SELECTION OF GABOR CONVOLUTION FEATURES[1]

2.1 Introduction

Face representation plays an important role in face recognition. Good face representation should be not only highly discriminative but also of low-dimensionality. In this chapter, we propose a new face representation method the uses non-uniformly selected Gabor convolution features.

Face recognition is a challenging research topic, since even the face of the same person can appear differently due to differences in lighting conditions, expression, pose, occlusion and other confounding factors [1]-[3]. To facilitate the face recognition task, it is important to accurately capture the local features in face images. Thus, a spatial-frequency analysis is advantageous. Wavelet analysis is useful for this purpose since it offers good spatial-frequency localization characteristics. In computer vision, the multi-resolution scheme in wavelet analysis has been justified by psycho-visual research [4].

Among various wavelet bases, Gabor wavelets provide a favorable tradeoff between spatial resolution and frequency resolution. There is a strong biological relevance for processing images using Gabor wavelets [5]. Gabor wavelets, whose kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells [6], have been proven to be capable of deriving desirable features in face recognition, mainly spatial frequency (scale), spatial locality, and orientation selectivity. The Gabor wavelet transform possesses useful properties such as invariance to illumination, rotation, scale, and translation. Previous

[1] A version of this chapter has been accepted for publication in IEEE Transactions on Systems, Man and Cybernetic, Part B, Shan Du and Rabab Ward, "Improved Face Representation by Non-uniform Multi-level Selection of Gabor Convolution Features."

research has demonstrated that using Gabor wavelets at the front-end of an automated face recognition system is effective [7]-[10].

The Gabor wavelet representation of an image involves convolution of the image with a family of Gabor kernels at different spatial frequencies and different orientations. To encompass the different spatial frequencies, spatial localities and orientation selectivities, the resulting convolution representations are normally concatenated as an augmented feature vector [11]. A face image is typically represented as the convolution result of the face image with 40 Gabor wavelets (5 scales, each with 8 orientations) [11][12]. Keeping only the magnitude values in the representation, this gives an '$h \times w \times 40$' vector, where $h \times w$ is the length of the face vector. Unfortunately, the '$h \times w \times 40$' vector usually has very large dimensionality.

To reduce the dimensionality of this vector, uniform sampling of the original Gabor features is traditionally used. The drawback of using uniform sampling is that it falsely assumes that all Gabor features contribute equally to the face recognition task. As a result, it may lead to a loss of features that are important while preserving many redundant ones. If the original Gabor features are finely sampled, the resulting vector will be still too large and will preserve many redundant or trivial features. The required system resources (i.e., CPU and memory) will be large and the processing speed will be slow. On the other hand, if the features are coarsely sampled, some important features may be lost and the recognition rate will be low.

We notice that different Gabor features contribute differently to the face recognition task. To overcome the problems of uniform sampling and improve the system's performance, we propose selecting the Gabor features according to their effectiveness in recognition, using a

statistical, non-uniform multi-level sampling procedure. The features corresponding to important face regions are sampled at a much finer rate than other parts of the image. The proposed method is based on the local statistics of the Gabor features, and is implemented using a coarse-to-fine hierarchical strategy. The sampled Gabor features are then classified using principal component analysis and/or linear discriminant analysis. Experiments conducted on representative benchmark face image databases show that the proposed method attains significantly higher recognition rates with much lower data dimensions than existing methods.

The remainder of this chapter is organized as follows: Section 2.2 summarizes the related work and discusses the problems inherent in uniform sampling of Gabor features. In Section 2.3, the proposed non-uniform multi-level selection approach is described in detail, and the differences from other methods are discussed. In Section 2.4, we evaluate the performance of the proposed method in face recognition based on experiments using four representative benchmark face databases. Performance improvements over existing methods are demonstrated in these experiments. Section 2.5 concludes the chapter.

2.2 Related Work

In recent years, Gabor wavelet-based face representation and recognition methods have been studied intensively [7], [13]-[19]. Lades *et al.* applied Gabor wavelets to face recognition via the dynamic link architecture (DLA) framework [13]. DLA first computes the Gabor *Jets*, and then performs a flexible template comparison among the resulting image decompositions using graph-matching. Wiskott *et al.* further expanded DLA and developed a Gabor wavelet-based elastic bunch graph matching method (EBGM) [7]. EBGM represents a face as a labeled graph. Each vertex of the graph corresponds to a manually predefined facial

landmark with fixed high-level semantics, labeled by multi-scale, multi-orientation Gabor Jets computed from the image area centered at the vertex landmark. An edge of the graph represents the connection between the two vertex landmarks and is labeled by the distance between them. After construction of the graph, identification can be achieved by elastic matching between the reference and probe graphs.

Another straightforward way to exploit Gabor features for face recognition was proposed by Liu [11]. In this method, multi-scale and multi-orientation Gabor features for each pixel in the normalized face images (with the eyes aligned) are computed and concatenated to form a high-dimensional Gabor feature vector. This vector is then uniformly down-sampled to form a low-dimensional feature vector, further reduced by principle component analysis (PCA), and then discriminated by enhanced Fisher discriminant analysis for final face identification. This method is simple and only needs to localize the two eyes as facial landmarks. However, the uniform down-sampling used in this method rejects a great number of informative Gabor features and preserves many redundant ones, which can unfavorably affect the final classification.

A non-uniform multi-level selection scheme is proposed here to select the more informative features and remove the redundant ones introduced by uniform sampling. The proposed non-uniform sampling is an efficient feature selection tool to reduce the dimensionality of Gabor features while improving the recognition rate. Our method is substantially different from existing Gabor feature selection schemes [14]-[19]. A brief yet complete survey of existing feature selection schemes is given below, as well as a comparison of each with our proposed scheme.

In [14], a learning algorithm that learns the importance of each uniformly sampled grid point for pose estimation task was employed. To select the most informative grid point, the jet response of each point is used as an isolated feature vector for a given test image. Using this feature vector only, the pose estimation performance of the system is calculated. A higher estimation performance means that the selected point contains useful information, and thus deserves a higher weight. By applying this scheme to each grid point on the sampling lattice, weights for each point are obtained.

[15] and [16] used the Adaboost method to assign weights for features. The authors used Adaboost to select a small set of Gabor features (or weak classifiers) from the original Gabor feature space to form a final strong classifier. The resulting classifier combines several hundred weak classifiers to calculate the similarity of a pair of Gabor faces. The Adaboost method needs to select the most informative Gabor features one by one from a large number of features.

In [17], the feature selection problem was formalized as a subset selection problem. A number of feature selection algorithms and a genetic algorithm were used to perform feature selection.

In [18], a graph-matching method was proposed. Instead of using predefined graph nodes as in the elastic graph matching (EGM) scheme, this method selects the peaks (high-energized points) of the Gabor wavelet responses to be used as fiducial feature points.

A relatively general method for attributing weights to features for a classification task was proposed in [19]. In this method, an object is represented as a labeled graph. The weighting is defined by a nonlinear function J that depends on a small set of parameters and on a training set. This function is the same for all graph nodes. The parameters are

determined on a training set by maximizing an evaluation function using the simplex method. Although quite general in theory, the application of this method requires some fine tuning and some a priori choices. The optimal settings and the particular choices seem to have been obtained by trials and testing.

The major differences between our algorithm and the above methods lie in its computational simplicity and its non-uniform sampling strategy. The method described in [14] is computationally demanding, as it examines the isolated performance of each uniformly sampled grid point to evaluate its importance. The training procedures in [15]-[17] are much more complex than that in our proposed method and thus require more computations. The method in [18] uses the high peak points as the elastic graph nodes and employs elastic graph matching. However, the high energy of the points does not necessitate that they contribute more information to the classification task. [18] and [19] are both related to the EBGM method, where the vertices of the graph are manually predefined to facial landmarks with fixed high-level semantics. The high complexity of accurate landmark predefining, graph construction and matching employed in EBGM limit its wide application. Our method differs from [18], [19] and EBGM [7] in that it does not need to predefine landmarks or perform graph construction, thus it requires significantly less computational time. An illustrated, detailed explanation of these differences is given in Section 2.3.

In the proposed method, the Gabor feature selection considers the local variance ratios of the Gabor features. Those features with high amplitude variance ratios are more informative than others. Therefore, the face regions with high variance ratios are sampled at higher sampling rates. This adaptation is implemented in a hierarchical fashion; a coarse-to-fine strategy results in multi-level sampling rates. The resulting non-uniformly sampled Gabor

features are then used for final classification. The proposed method has the advantages of low complexity, low dimensionality and of being highly discriminative.

In the following sections, a detailed description and performance evaluation of the proposed method are given.

2.3 Gabor Feature Selection using Multi-level Non-uniform Sampling

2.3.1 Gabor Wavelets and Gabor Features

Gabor wavelets extract the information quanta in forms of space and frequency, two physically measurable quantities, combined in the most elegant way by Heisenberg's uncertainty relation [3]. It is well known that Gabor wavelets effectively model the receptive field profiles of cortical simple cells in the primary visual cortex [6]. The Gabor wavelet representation, therefore, captures the salient visual properties such as spatial localization, orientation selectivity, and spatial frequency.

The 2-D Gabor wavelets (kernels, filters) can be defined as follows:

$$\psi_j(\vec{x}) = \frac{\|\vec{k}_j\|^2}{\sigma^2} \exp\left(-\frac{\|\vec{k}_j\|^2 \|\vec{x}\|^2}{2\sigma^2}\right)\left[\exp\left(i\vec{k}_j\vec{x}\right) - \exp\left(-\frac{\sigma^2}{2}\right)\right], \quad j = 1 \to L \qquad (2\text{-}1)$$

where $\vec{k}_j = k_m e^{i\varphi_n}$ (2-2)

$$k_m = \frac{0.5\pi}{\left(\sqrt{2}\right)^m} \quad \varphi_n = n\frac{\pi}{8} \qquad (2\text{-}3)$$

$$L = m \times n \qquad (2\text{-}4)$$

m and n define the scale and orientation of the Gabor wavelets, $\|\ \|$ denotes the norm operation, and $\vec{x}$ represents the pixel position.

In most cases, the Gabor wavelets at five different scales ($m \in \{0,...,4\}$), and eight

orientations ($n \in \{0,...,7\}$) are used. Figure 2-1 shows all 40 Gabor wavelets (at different scales and orientations) used in our proposed method.

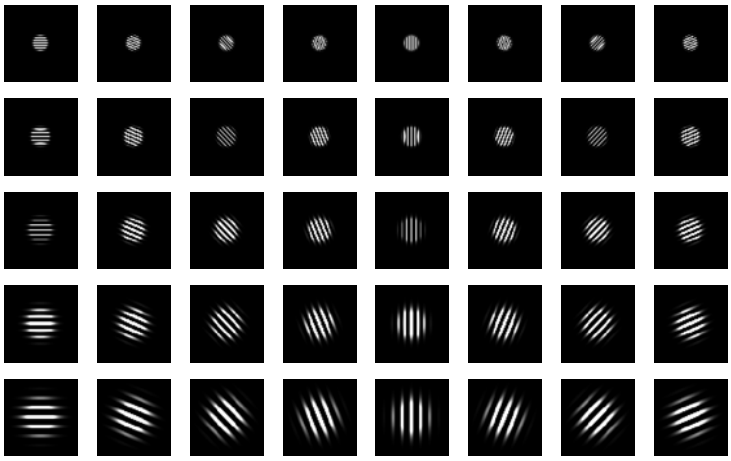


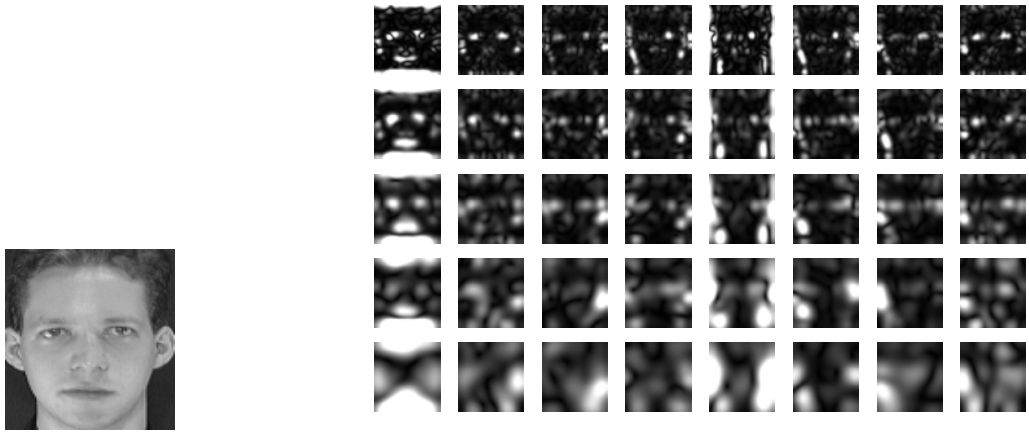
Figure 2-1. The 40 Gabor wavelets used in the proposed method, at five scales and eight orientations.

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor wavelets. Let $I(\vec{x})$ be the $h \times w$ gray-level image, $\psi_j(\vec{x})$ be the j^{th} Gabor wavelet, and $O_j(\vec{x})$ be the j^{th} Gabor feature image of image $I(\vec{x})$ corresponding to the j^{th} Gabor wavelet $\psi_j(\vec{x})$. Then

$$O_j(\vec{x}) = I(\vec{x}) * \psi_j(\vec{x}) \qquad j = 1 \to L \tag{2-5}$$

where $(\vec{x}) = (x, y)$, * denotes the convolution operator, and $L = m \times n$ is the number of Gabor wavelets. Here, $L = 40$.

The Gabor feature images of an example image, from the ORL face image database, are shown in Figure 2-2. The feature images exhibit strong characteristics of spatial locality, scale, and orientation selectivity corresponding to those displayed by the Gabor wavelets in Figure 2-2. Such characteristics produce salient local features that are suitable for visual recognition.



a. An ORL database face image b. Magnitude of the Gabor features

Figure 2-2. Gabor features of an example image.

2.3.2 Non-uniform Multi-level Selection of Gabor Features

The proposed face representation method is now described. Figure 2-3 shows a block diagram of the entire face representation and recognition system. First, the Gabor wavelet transform removes most variability in the images resulting from variations in lighting and contrast. Then the Gabor features $O_j(\vec{x})$ are selected based on a statistical study of a set of training face images. The ratio of the between-class variance to the within-class variance is computed for each feature extracted from the training set. The features with high variance ratios are considered to be more informative than those with low variance ratios. The face regions corresponding to high variance ratios are assigned higher sampling rates. The non-uniformly selected Gabor features (referred to as *NUGF*) are weighted according to a simple criterion and then further reduced in dimensionality by PCA. They are then fed into the linear discriminant analysis classifier. The classification scheme employs the Euclidean distance and conventional nearest neighbor method for implementation simplicity.

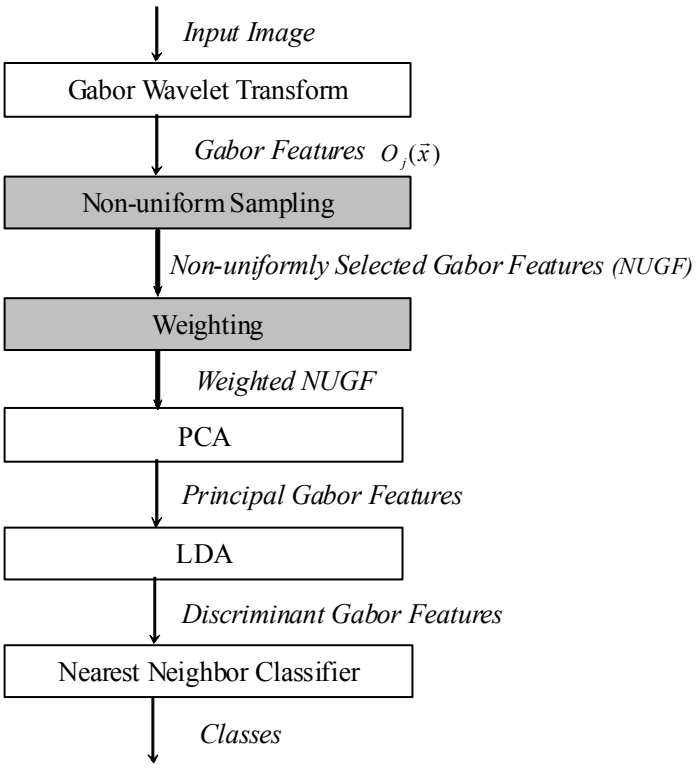


Figure 2-3. Block diagram of the system (the grey blocks are the major contributions).

Since the Gabor feature images $O_j(\vec{x}),\ j=1\to L$ consist of different local, scale, and orientation features, we concatenate all $O_j(\vec{x}),\ j=1\to L$, and derive a feature vector, X. Since the image size is $h\times w$, and we apply 40 Gabor wavelets, the dimension of the derived feature vector, X, is $h\times w\times 40$. As the resulting vector is too long, each $O_j(\vec{x})$ is normally down-sampled, before concatenation, by a factor ρ, e.g., $\rho=8\times 8$. This last step is traditionally performed using uniform sampling.

The choice of the sampling rate used for uniform sampling is a tradeoff between vector dimensionality and retention of important features. Fine down-sampling, e.g., by $\rho=2\times 2$, results in a large vector but preserves many redundant features. Coarse sampling, e.g., $\rho=32\times 32$, may result in the loss of some important features. We note that the $h\times w\times 40$ features do not contribute equally to the face recognition task. Since uniform sampling treats

all features equally, it will not result in optimal performance. Thus, a non-uniform sampling approach is proposed here that assigns higher importance to the more relevant Gabor features by sampling them at finer rates, and uses sparse samples for the less relevant features.

To define the importance of the different Gabor features, a total of *K* training images of *c* classes are used. The Gabor features for each training image are calculated via Gabor transform. The ratios of the between-class variance to the within-class variance of the Gabor features of the training set are obtained using Equations (2-6) and (2-7). To find the more important features, i.e., the ones with more information, we notice that these features have higher variance ratios than the others. This method can also work when only one training image is available for each person. In this case, the variance ratio is computed using Equation (2-8).

Let $\sigma_j^{\,2}(\vec{x})$ be the $h \times w$ matrix of the ratios of the between-class variance to the within-class variance of the Gabor feature image $O_j(\vec{x})$. Each specific entry (x, y) in $\sigma_j^{\,2}(\vec{x})$ corresponds to the variance ratio of $O_j(x, y)$. To calculate $\sigma_j^{\,2}(\vec{x})$, we use a set of *K* training images of *c* classes (i.e., *c* different persons) $\{X_1, X_2, ..., X_c\}$, where each class has K_i images and $K = \sum_{i=1}^{c} K_i$. The Gabor feature images of all *K* training images are $O_j^{\,k}(\vec{x})$, $j = 1 \to L$, $k = 1 \to K$. The Gabor feature images of the K_i training images of class X_i are $O_j^{\,k_i}(\vec{x})$, $j = 1 \to L$, $k_i = 1 \to K_i$.

The between-class variance of Gabor features $O_j(\vec{x})$ is defined as

$$\sigma_{Bj}^{\,2}(\vec{x}) = \frac{\sum_{i=1}^{c} K_i [\mu_{ij}(\vec{x}) - \mu_j(\vec{x})]^2}{K} \qquad (2\text{-}6)$$

where $\vec{x}$ is the pixel position; $\mu_{ij}(\vec{x})$ is the mean of the K_i Gabor feature images $O_j^{k_i}(\vec{x})$, $k_i = 1 \to K_i$ of class X_i; $\mu_j(\vec{x})$ is the mean of the K Gabor feature images $O_j^{k}(\vec{x})$, $k = 1 \to K$ of all classes; K is the number of training images and K_i is the number of samples in class X_i; c is the number of classes; $\mu_{ij}(\vec{x}) = \frac{\sum_{k_i=1}^{K_i} O_j^{k_i}(\vec{x})}{K_i}$;

$$\mu_j(\vec{x}) = \frac{\sum_{k=1}^{K} O_j^{k}(\vec{x})}{K} = \frac{\sum_{i=1}^{c}\sum_{k_i=1}^{K_i} O_j^{k_i}(\vec{x})}{\sum_{i=1}^{c} K_i}.$$

The within-class variance of Gabor features $O_j(\vec{x})$ is defined as

$$\sigma_{Wj}^{\ 2}(\vec{x}) = \frac{\sum_{i=1}^{c} \sum_{k_i=1(O_j^{k_i}(\vec{x}) \in X_i)}^{K_i} [O_j^{k_i}(\vec{x}) - \mu_{ij}(\vec{x})]^2}{K} \qquad (2\text{-}7)$$

The variance ratio $\sigma_j^{\ 2}(\vec{x})$ of Gabor features $O_j(\vec{x})$ is then given by $\frac{\sigma_{Bj}^{\ 2}(\vec{x})}{\sigma_{Wj}^{\ 2}(\vec{x})}$.

When only one sample image is available for each person, only the between-class variance applies. The variance ratio is considered to be equal to $\sigma_{Bj}^{\ 2}(\vec{x})$. Thus, Equation (2-6) needs to be transformed from $\sigma_{Bj}^{\ 2}(\vec{x}) = \frac{\sum_{i=1}^{c} K_i [\mu_{ij}(\vec{x}) - \mu_j(\vec{x})]^2}{K}$, where $K_i = 1, \quad i = 1...c; c = K; \ \mu_{ij}(\vec{x}) = O_j^{\ k}(\vec{x})$ to

$$\sigma_j^{\ 2}(\vec{x}) = \frac{\sum_{k=1}^{K} \left[O_j^{\ k}(\vec{x}) - \mu_j(\vec{x}) \right]^2}{K} \qquad (2\text{-}8)$$

The non-uniform sampling rates are assigned via a coarse-to-fine sampling strategy. In this strategy, a Gabor feature image $O_j(\vec{x})$ is first partitioned into many regions by a coarse grid, e.g., $\rho = 16 \times 16$. Then each region of size 16×16 features is considered separately, e.g., the region ABCD in Figure 2-4. The local mean variance $\overline{\sigma}_j^{2}{}_{local}$ of this region is calculated. If the local mean variance is not higher than the global mean variance $\overline{\sigma}_j^{2}{}_{global}$, measured on all Gabor feature images, then this region is represented by one point only. Otherwise, this region is considered important, and is further divided into 4 sub-regions, e.g., AEIH, EBFI, HIGD, and IFCG in Figure 2-4. For each newly generated sub-region, the above procedure is repeated. If the local mean variance of a sub-region is found to be higher than the global mean variance, then that sub-region is further sub-divided. This procedure is repeated until the sub-region reaches a certain predefined size, e.g., 4×4. The average value of the Gabor features of each region is selected to represent that region. Figure 2-4 illustrates the process. Figure 2-5c shows the samples corresponding to the above Gabor feature sampling strategy.

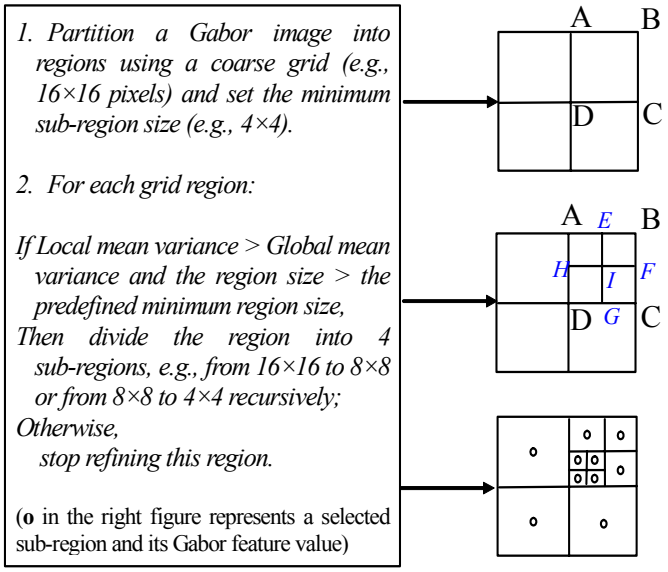
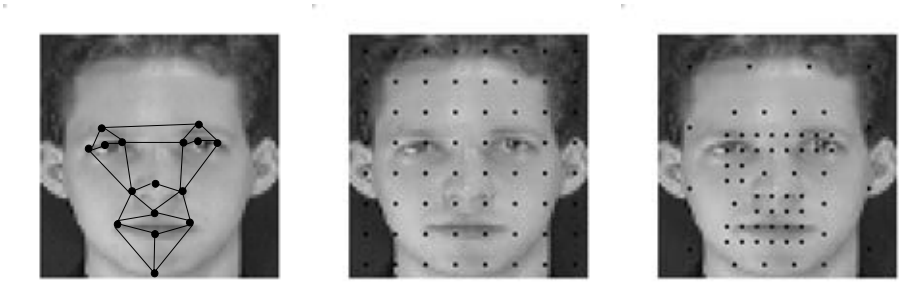


Figure 2-4. Non-uniform sampling algorithm.

Figure 2-5 illustrates the differences between the proposed method and uniform sampling and EBGM. Because non-uniform sampling saves some bits on the less relevant Gabor features, we can use a finer sampling rate for the more relevant features. When uniform and non-uniform samplings have the same number of samples, they produce a significant performance difference. Non-uniform sampling is implemented automatically without facial landmark position initialization as in EBGM (Figure 2-5a), where facial landmarks are first manually predefined with canonical positions that are subsequently shifted and adapted to the input face geometry and all the 40 Gabor features (5 scales and 8 orientations) are computed for each landmark. In uniform sampling, all the 40 Gabor filters are convolved with the image at each vertex of a uniform grid. In our method, the Gabor filters (in terms of its position, the scale and the orientation) are learned automatically and obtained by the non-uniform sampling strategy.



a. EBGM b. uniform c. non-uniform

Figure 2-5. Differences between EBGM, uniform sampling and the proposed method.

2.3.3 Feature Weighting Strategy

When the selection of the Gabor features is completed, a weight is assigned to each selected feature. This is because even after non-uniform sampling, some Gabor features remain more important for recognition purposes than others. To simplify the explanation of the weighting procedure, an example is given. Assume the initial and the smallest sampling grid sizes are chosen to be 16×16 and 4×4, respectively. Thus, the possible sizes of a sub-

region for which one sample is assigned are 16×16, 8×8 or 4×4. If a sub-region of size 16×16 (such as 1, 2, and 3 in Figure 2-6a) is not sub-divided into smaller sub-regions, then the sample point representing it is assigned the lowest weight of 1 (denoted as *grey* points). For the 8×8 sub-regions, such as 4, 5, and 6 in Figure 2-6b, which are not further sub-divided, we assign a higher weight of 2 (denoted as *black* points) to their corresponding sample points. As for the sub-regions of size 4×4, they are of two kinds. For example, sub-regions 9 and 10 in Figure 2-6c have higher local variances than the global variance of the image. Thus, if the size of the smallest sub-region was set as 2×2 instead of 4×4, then more samples would have been generated. Therefore, we assign to their sample points the highest weight of 3 (denoted as *white* points) to emphasize their importance. As for the sub-regions 7 and 8, their local variances are lower than the global variance. These sub-regions would not have been further sub-divided even if the size of the smallest sub-region was reduced. In order not to make the weighting process too complex, we assign their sample points a weight of 2 (denoted as *black* points), i.e., the same weight assigned to sub-regions 4, 5, and 6.

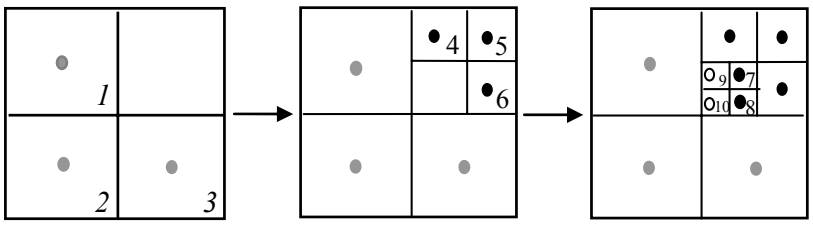


Figure 2-6. The coarse-to-fine selection process and the weighting process.

In summary, we assign the highest weight of 3 to the samples whose corresponding sub-regions would have been further sub-sampled had they not been constrained by the imposed size of the smallest sub-region. We assign the lowest weight of 1 to the samples with the initial minimum sampling rate, and a weight of 2 to the remaining samples. In Figure 2-7, the

grey points indicate the samples with weight 1, the black points indicate those with weight 2, and white points with weight 3. Note that this weighting strategy does not require any extra computations. The weights are obtained during the statistical selection process, yet their use improves recognition performance with no additional computational burden.

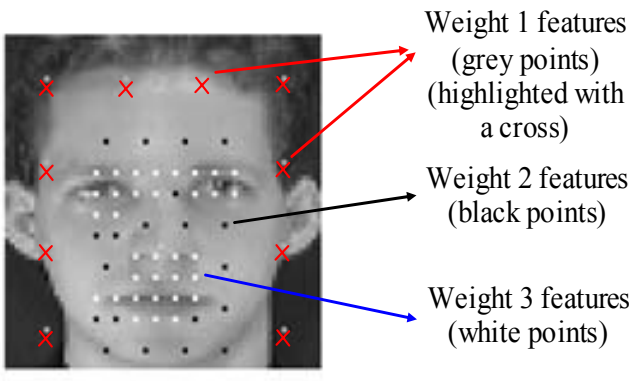


Figure 2-7. Weighted non-uniformly sampled Gabor features shown on a face.

2.3.4 Linear Discriminant Analysis and Principal Component Analysis of the Weighted NUGFs

Linear discriminant analysis (LDA) is recognized as a useful analysis and classification tool [20] that transforms the original face representation to a new sub-space, where the between-class scatter is maximized and the within-class scatter is minimized. In this chapter, we apply an LDA classifier to the weighted non-uniform Gabor features (NUGFs) for classification purposes. Let $f_1, f_2, ..., f_K$ represent the K vectors of the weighted NUGFs of a training set composed of K face images belonging to one of c classes $\{X_1, X_2, ..., X_c\}$, where each class has K_i images. Vector f_j is of length N.

We define two measures: (1) the within-class scatter matrix, given by

$$S_w = \sum_{i=1}^{c} \sum_{j=1}^{K_i} (f_j^i - \bar{f}^i)(f_j^i - \bar{f}^i)^T \qquad (2\text{-}9)$$

where f_j^i is the *j*th sample of class X_i, $\bar{f}^i$ is the mean of class X_i; and (2) the between-class scatter matrix

$$S_b = \sum_{i=1}^{c} K_i (\bar{f}_j^{\,i} - \bar{f})(\bar{f}_j^{\,i} - \bar{f})^T \tag{2-10}$$

where $\bar{f}$ represents the mean of all classes.

The goal is to maximize the between-class measure while minimizing the within-class measure. One way to do this is to maximize the ratio $\frac{\det|S_b|}{\det|S_w|}$. If S_w is a nonsingular matrix, then this ratio is maximized when the column vectors of the projection matrix are the eigenvectors of $S_w^{-1}S_b$. It should be noted that S_w may be singular. To solve this singular problem, [20] proposed using the principal component analysis (PCA) space as an intermediate space. PCA is first conducted to further reduce the dimensionality of the NUGFs to be less than *K*-*c*, where *K* is the number of training examples and *c* is the number of classes. Thus, the original *N*-dimensional space is projected onto an intermediate *M*-dimensional space using PCA and then onto a final *P*-dimensional space using LDA.

PCA is a standard technique used to approximate the original data by lower dimensional feature vectors [21]. It transforms interdependent coordinates into significant and independent ones. Considering the weighted NUGFs $f_1, f_2, ..., f_K$ whose average is $\bar{f}$, the sample matrix representing the *K* weighted NUGF vectors is the following *N* by *K* matrix:

$$A = \begin{bmatrix} f_1'(1) & f_2'(1) & \dots & f_K'(1) \\ f_1'(2) & f_2'(2) & \dots & f_K'(2) \\ \dots & \dots & \dots & \dots \\ f_1'(N) & f_2'(N) & \dots & f_K'(N) \end{bmatrix} \tag{2-11}$$

where $f_i' = f_i - \bar{f}$.

The covariance matrix is defined as

$$C = \sum_{i=1}^{K} f_i' f_i'^T = AA^T \qquad (2\text{-}12)$$

The eigenvalues and the eigenvectors of the covariance matrix (2-12) are then obtained and the weighted NUGFs of an input image is transformed into its principal components by convolving it with the eigenvectors.

The convolution coefficients form a vector $\Omega = [\omega_1, \omega_2, ..., \omega_K]$ that describes the contribution of each eigenvector in representing the input weighted NUGFs, treating the eigenvectors as a basis set for the weighted NUGFs:

$$\omega_k = \alpha_i^T (f_{in} - \bar{f}), \quad i = 1,2,...,K \qquad (2\text{-}13)$$

where α_i is the set of eigenvectors of C and f_{in} is the weighted NUGFs of the input image. These coefficients ω_k are then fed into the LDA classifier for the final classification.

Note that LDA can only work when there are multiple sample images available for each person in the training set [23]. By contrast, our method does not need multiple images for each person. To address the case when only one sample image per person is available, Equation (2-10) is transformed from

$$S_b = \sum_{i=1}^{c} K_i (\mu_i - \mu)(\mu_i - \mu)^T, \text{ where } K_i = 1, \quad i = 1...c;\ c = K;\ \mu_i = f_i;\ f_i' = f_i - \mu$$

to

$$S_b = \sum_{i=1}^{K} f_i' f_i'^T = C \qquad (2\text{-}14)$$

Since S_w does not exist, maximizing the ratio $\frac{\det |S_b|}{\det |S_w|}$ is maximizing C, which is actually the PCA.

2.4 Experiments and Performance Analysis

The advantages of the proposed method are as follows: (1) there is no manual facial landmark position initialization, (2) there is no model construction or fitting which is time consuming, (3) the training is very simple, (4) it has a low dimensional data and achieves a high recognition rate, and (5) it works well even when there is only one training image available for each person.

To evaluate the effectiveness of the proposed method, face images with popular imaging variations, such as expression, illumination, image scale and pose are used. The distribution of the selected Gabor features, the recognition rates and the dimensionality of features are analyzed. The effectiveness of the proposed method is demonstrated by comparing its results with those of popular methods, such as the eigenface method (PCA) [21], Fisherface method (LDA) [20], and the typical Gabor features-based methods, such as the combination of uniform Gabor and eigenface method, the combination of uniform Gabor and Fisherface method, the Gabor-Fisher classifier (GFC) [11], the Adaboost Gabor Fisher classifier (AGFC) [16], and a method proposed in [18].

2.4.1 Testing Face Databases

Experiments are performed on four widely-used benchmark face databases: (1) Olivetti Research Laboratory (ORL), (2) Yale, (3) Yale B, including the extended database, and (4) Facial Recognition Technology (FERET) database.

The ORL database [24] is composed of 400 images with ten different images for each of the 40 distinct subjects. The images vary across pose, size, time, and facial expression. Figure 2-8 shows some images that we use from the ORL face database. They are resized to 64×64.

Figure 2-8. Examples of face images in the ORL database.

The Yale database [20][25] contains 165 grayscale images of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised, and winking. The Yale database images have a fixed pose, but different illumination conditions and expressions. Examples of the normalized Yale images are shown in Figure 2-9. They are resized to 64×64.

Figure 2-9. Examples of the Yale images used in our experiments.

The Yale Face Database B [26] contains 5760 single light source images of 10 individuals. Each subject has 9 poses and each pose has 64 different illumination conditions. The extended Yale Face Database B [27] contains 16128 images of 28 human subjects under 9 poses and 64 illumination conditions. We take a subgroup of the original and the extended Yale B database images that have fixed illumination and different poses. The range of poses is from the half left-side profile to half right-side profile. Figure 2-10 shows the 10 subjects

used from the original Yale Face Database B and Figure 2-11 shows the 28 subjects used from the extended Yale Face Database B. Example images of one person with front-lit illumination in 9 poses are shown in Figure 2-12, resized to 64×64.

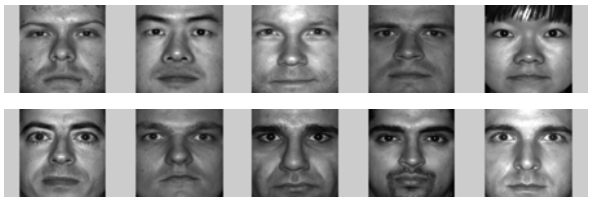


Figure 2-10. 10 subjects of the original Yale Face Database B.

Figure 2-11. 28 subjects in the extended Yale Face Database B.

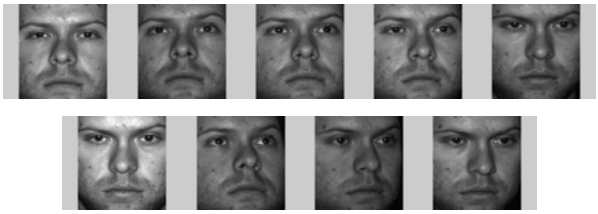


Figure 2-12. 9 different poses of each person.

The Facial Recognition Technology (FERET) database [29][30] displays diversity across gender, ethnicity, and age. It consists of 14051 images and has a strictly separate testing set (composed of Gallery and Probe sets) and a training set. Figure 2-13 shows some example images used in our experiments, cropped to size 64×64. Note that the images were acquired during different photo sessions under variable illuminations and facial expressions.

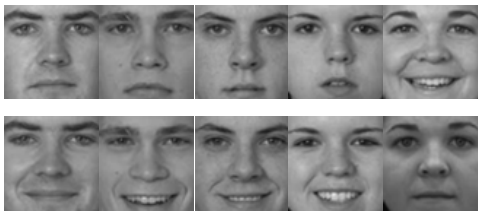


Figure 2-13. Example FERET images used in our experiments.

2.4.2 Analysis of the Non-uniformly Selected Gabor Features (NUGFs)

As mentioned earlier, the non-uniformly selected Gabor features should be the most informative features to discriminate different faces. To observe the characteristics of these features, we conduct experiments on the ORL training sets consisting of five images per person, to obtain NUGFs. Some of the statistics of these NUGFs are given below.

2.4.2.1 Position distribution of the NUGFs

Figure 2-14 shows the discriminative NUGFs obtained by non-uniform sampling. From the figure, we can easily see that the eyes, nose and mouth provide more information than other parts for face recognition.

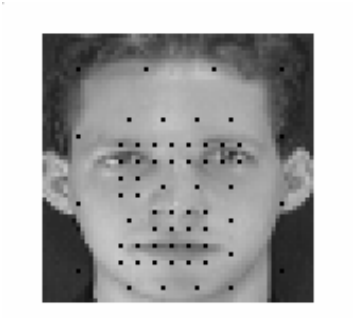


Figure 2-14. Positions of the discriminative Gabor features in a face.

2.4.2.2 Distribution of the 40 Gabor kernels in the NUGFs

Figure 2-15 illustrates the occurrence frequency of the 40 Gabor kernels in the NUGFs. From the figure, we can conclude that different Gabor kernels contribute in different ways to

face representation and recognition. For the ORL database, the Gabor kernels 1 (m=0, n=0), 8 (m=0, n=7), 9 (m=1, n=0), and 17 (m=2, n=0) contribute more than the others.

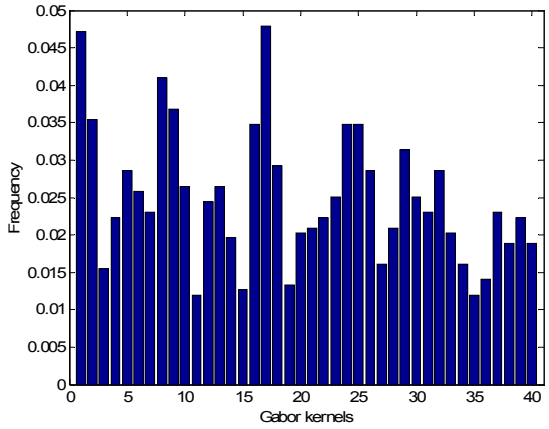


Figure 2-15. Distribution of the 40 Gabor kernels in the NUGFs.

2.4.2.3 Scale distribution of the NUGFs

Figure 2-16 illustrates the distribution of Gabor scales 0→4 in the NUGFs. From this figure, we can see that kernels with 0-scale contribute more to face representation and recognition than the others do.

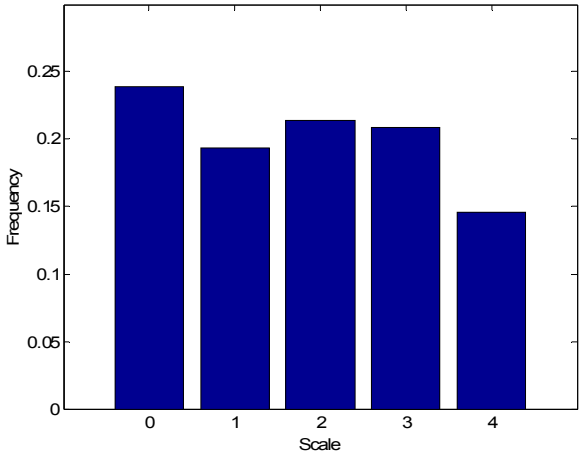


Figure 2-16. Scale distribution of the NUGFs.

2.4.2.4 Orientation distribution of the NUGFs

Different orientations also contribute differently to face representation and recognition. Figure 2-17 illustrates the distribution of orientations 0→7 in the NUGFs. From this figure, we can see that orientations 0, 1, and 7 contribute more than the others do.

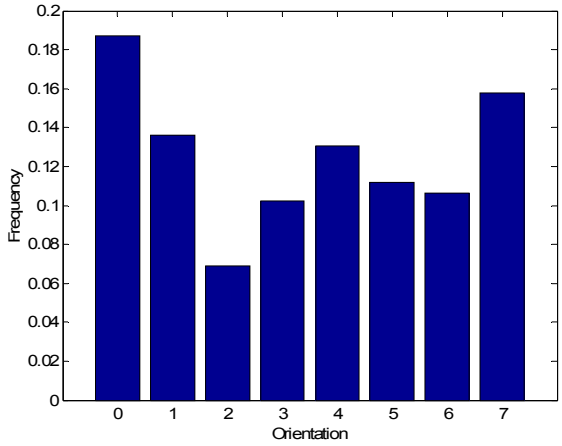


Figure 2-17. Orientation distribution of the NUGFs.

2.4.3 Performance Analysis

In this section, we show the data dimension and recognition rate of the proposed method. First, we implement four representative methods, eigenface method (PCA), Fisherface method (LDA), the combined uniform Gabor and eigenface method (uniform-Gabor-based eigenface), and the combined uniform Gabor and Fisherface method (uniform-Gabor-based Fisherface), to compare with the proposed new method, where non-uniform-Gabor-based eigenface and non-uniform-Gabor-based Fisherface are both compared.

The experiments are divided into two cases. In Case 1, only one training image is available for each person. One image for each person is randomly chosen for training and the remaining images are used for testing (for ORL, Yale and Yale B database). In Case 2, multiple sample images are available for each person. For the ORL database, five images are

randomly chosen from the ten images available for each person for training, and the remaining five images are used for testing. For the Yale database, five images are randomly chosen for training, while the remaining six images are used for testing. For the Yale B database, four images in four different poses of each of the 38 persons are used as the training images. The remaining five images in five other different poses of each person are used for testing.

For Case 1, because only one sample image is available for each person, only PCA is applied to classification. For Case 2, both PCA and LDA-based classification are implemented.

Second, in order to facilitate the comparison of our method with other Gabor feature selection methods, the FERET database is employed. Note that the FERET database has strictly distinguished the testing set (composed of Gallery and Probe sets) from the training set. We test our method on the largest probe set FB with 1195 images of different subjects. There are 1196 images in the gallery set. The training set we use is a near-frontal face subset of the standard FERET training set, in which only the near-frontal face images in the standard FERET training CD are included. Table 2-1 shows the structure of the FERET face database we use to evaluate our method.

Table 2-1. Structure of the FERET face database used in our experiments.

Database		**No. of Persons**	**No. of Images**	**Note**
Training Set		431	1000	All near-frontal faces in the standard FERET training set
Test Set	Gallery	1196	1196	Standard FERET gallery with near-frontal faces
	FB Probes	1195	1195	Near-frontal faces with different expressions from those in gallery

2.4.3.1 Comparison of data dimensionality of the selected features

All the methods in this chapter use PCA and/or LDA for classification. The data dimensionalities of the inputs to PCA/LDA are compared. The lower the dimensionality, the faster the PCA/LDA is computed. For the eigenface and Fisherface methods, the inputs are the face images; for the others, the inputs are the sampled Gabor features (either uniform or non-uniform).

Let w and h denote the width and the height of a face image respectively. For the eigenface/Fisherface methods, the dimension of the inputs to PCA/LDA is

$$N_{orig} = w \times h \tag{2-15}$$

For uniform sampling, let r denote the sampling rate. The dimension of the inputs to PCA/LDA is

$$N_{unif} = (\frac{w}{r}) \times (\frac{h}{r}) \times 40 \tag{2-16}$$

For the non-uniform sampling, let r_i denotes the initial sampling rate and r_m denotes the maximum sampling rate. The dimension of the inputs to PCA/LDA is between $N_{non-unif-\max}$ and $N_{non-unif-\min}$, which are given by

$$N_{non-unif-\max} = (\frac{w}{r_m}) \times (\frac{h}{r_m}) \times 40 \tag{2-17}$$

$$N_{non-unif-\min} = (\frac{w}{r_i}) \times (\frac{h}{r_i}) \times 40 \tag{2-18}$$

The actual data dimension in the non-uniform sampling depends on the specific database.

In the experiments, the size of a face image is 64×64. Therefore, for the eigenface and Fisherface methods, the data dimension is 4096. For the uniform sampling based methods, the uniform sampling rate of Gabor features is 64. Thus, the data dimension after uniform

sampling is 2560. For the newly proposed method, the most important Gabor features are selected by the non-uniform sampling procedure from the original 163840 Gabor features. Both the dimensionality of the Gabor features and the recognition rate are mainly controlled by the maximum sampling rate r_m . $r_i = 32$ (initial sampling rate) and $r_m = 4$ are chosen because of the trade-off between the recognition rate and the data size.

Table 2-2 to Table 2-8 show the comparisons of the data dimensions for the different methods. In contrast to the high dimensions in existing methods (4096 for eigenface and Fisherface; 2560 for uniform sampling), the data dimension in the new method is only about 700. The proposed method reduces the data size to less than *30%* of that using uniform sampling, and to less than *20%* of that using the original image. Compared with other Gabor features selection methods, our method still has much lower dimensions. Note that the dimensionality of the Gabor features influences the classification and recognition processing speed and the storage space. Feature vectors of lower dimensional spaces reduce the computation complexity (i.e., increase the speed), as well as the storage required for face recognition.

2.4.3.2 Comparison of recognition rates

Table 2-2 to Table 2-8 also give the recognition rates of different methods: eigenface (PCA), Fisherface (LDA), uniform-Gabor-based eigenface, uniform-Gabor-based Fisherface, non-uniform-Gabor-based eigenface and non-uniform-Gabor-based Fisherface.

From the experimental results, it can be seen that with much lower dimensionality (less than *30%* of the others), the non-uniformly sampled Gabor features method achieves significant improvement in recognition rate. The experimental results lead to the following findings:

1) The non-uniform multi-level Gabor feature selection method obtains higher recognition rates for any of the benchmark face databases. For example, for the Yale face database, the non-uniform Gabor features based Fisherface method improves the recognition rate from 91.1% (the original Fisherface method) and 93.3% (the uniform Gabor based Fisherface method) to 98.5%.

2) The uniformly sampled Gabor features of 2560-dimension produces better recognition results than the original eigenface/Fisherface methods with 4096-dimension data. This implies that the Gabor convolution features are more informative than the gray-scale pixel value.

3) The non-uniformly sampled Gabor features of about 700-dimension produces better recognition rates than the uniformly sampled Gabor features of 2560-dimension. This is because non-uniform sampling intelligently extracts more features on more discriminative face regions and ignores some trivial regions.

4) The Yale B database is the most difficult in the testing databases due to the large pose variations; however, our method is still satisfactory. It is possible to further improve the performance by a pre-processing procedure for pose variations. For example, using the 3D morphable model method [31][32] or the linear object classes method [33] to normalize a non-frontal image to a frontal image. Moreover, we believe our NUGF method can be extended to a view-based method.

5) In all methods, using multiple sample images for training generates better recognition rates than using a single image; however, our method produces satisfactory results even for the case with only one sample image.

6) The combination of non-uniform Gabor features and LDA yields the best recognition performance.

In order to compare our method with other Gabor feature selection methods, Table 2-8 also shows the comparison of the proposed method with the referenced Gabor feature selection methods [11][16][18] using the FERET database. The results show that the proposed method outperforms other methods in terms of both recognition rate and data dimension.

In summary, from the experimental results on both data dimensions and recognition rates, the overall performance is significantly improved with the new low complexity, low dimensional and highly discriminative face representation method.

Table 2-2. Performance comparisons on ORL database (one sample image per person).

Methods	**Data Dimensions**	**Recognition Rate**
eigenface (PCA)	4096	75.5%
uniform-Gabor-based eigenface	2560	80.8%
non-uniform-Gabor-based eigenface	706	86.2%

Table 2-3. Performance comparisons on ORL database (multiple sample images per person).

Methods	**Data Dimensions**	**Recognition Rate**
eigenface (PCA)	4096	84.9%
Fisherface (LDA)	4096	88.8%
uniform-Gabor-based eigenface	2560	95.3%
uniform-Gabor-based Fisherface	2560	95.9%
non-uniform-Gabor-based eigenface	676	98.1%
non-uniform-Gabor-based Fisherface	676	98.9%

Table 2-4. Performance comparisons on Yale database (one sample image per person).

Methods	Data Dimensions	Recognition Rate
eigenface (PCA)	4096	63.8%
uniform-Gabor-based eigenface	2560	79.3%
non-uniform-Gabor-based eigenface	673	87.1%

Table 2-5. Performance comparisons on Yale database (multiple sample images per person).

Methods	Data Dimensions	Recognition Rate
eigenface (PCA)	4096	88.9%
Fisherface (LDA)	4096	91.1%
uniform-Gabor-based eigenface	2560	92.2%
uniform-Gabor-based Fisherface	2560	93.3%
non-uniform-Gabor-based eigenface	703	96.4%
non-uniform-Gabor-based Fisherface	703	98.5%

Table 2-6. Performance comparisons on Yale B database (one sample image per person).

Methods	Data Dimensions	Recognition Rate
eigenface (PCA)	4096	45.4%
uniform-Gabor-based eigenface	2560	63.8%
non-uniform-Gabor-based eigenface	673	73.7%

Table 2-7. Performance comparisons on Yale B database (multiple sample images per person).

Methods	Data Dimensions	Recognition Rate
eigenface (PCA)	4096	74.7%
Fisherface (LDA)	4096	78.7%
uniform-Gabor-based eigenface	2560	80.9%
uniform-Gabor-based Fisherface	2560	83.9%
non-uniform-Gabor-based eigenface	784	86.6%
non-uniform-Gabor-based Fisherface	784	89.5%

Table 2-8. Performance comparisons of different Gabor feature selection methods on FERET database.

Methods	Data Dimensions	Recognition Rate
eigenface (PCA) [21]	4096	80.0%
Fisherface (LDA) [20]	4096	85.4%
uniform-Gabor-based eigenface	2560	90.2%
uniform-Gabor-based Fisherface	2560	94.4%
GFC (LDA) [11] [16]	9000	96.3%
Method in [18]	~1300	96.3%
AGFC (Adaboost) (LDA) [16]	1884	97.2%
non-uniform-Gabor-based eigenface	710	95.5%
non-uniform-Gabor-based Fisherface	710	97.2%

2.5 Conclusions

In this chapter, an efficient method for face representation is presented and its effectiveness demonstrated. This method relies on non-uniform multi-level sampling of the Gabor feature vectors. The new sampling strategy is based on using more samples for the more relevant parts and fewer samples for the less relevant ones. Sampling is performed using a coarse-to-fine hierarchical strategy with multi-level sampling rates. The sampling rate adaptation is implemented automatically without facial landmark position initialization. Experiments are conducted on the ORL, Yale, Yale B, and FERET face image databases, where the images vary in illumination, expression, pose, and scale. The results show that the proposed non-uniform sampling of the Gabor features outperforms the other face recognition methods. This method works well even when only one training image is available for each person. Besides its advantage in yielding significantly higher recognition accuracy, the proposed method greatly reduces the dimensionality of the features for classification and thus is computationally less demanding. By using the new non-uniform multi-level face representation method, the overall system performance is substantially improved.

2.6 References

[1] R. Chellappa, C. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proc. of the IEEE*, vol. 83, no. 5, pp. 705-740, 1995.

[2] A. Samal and P. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Pattern Recognition*, vol. 25, pp. 65-77, 1992.

[3] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, 2003.

[4] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America*, vol. 2, no. 7, pp. 1160-1169, 1985.

[5] J. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision Research*, vol. 20, pp. 847-856, 1980.

[6] D. Field, "Relations between the statistics of natural images and the response properties of cortical cell," *Journal of the Optical Society of America A*, vol. 4, no. 12, pp. 2379-2394, 1987.

[7] L. Wiskott, J. Fellous, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 775-779, 1997.

[8] J. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 36, pp. 1169-1179, 1988.

[9] T. Lee, "Image representation using 2D Gabor wavelets," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 959-971, 1996.

[10] C. Liu and H. Wechsler, "Independent component analysis of Gabor features for face recognition," *IEEE Trans. on Neural Networks*, vol. 14, no. 4, pp. 919-928, 2003.

[11] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. on Image Processing*, vol. 11, no. 4, pp. 467-476, 2002.

[12] C. Liu, "Gabor-based kernel PCA with fractional polynomial models for face recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 572-581, 2004.

[13] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. on Computers*, vol. 42, pp. 300-311, 1993.

[14] B. Gokberk, L. Akarun, and E. Alpaydin, "Feature selection for pose invariant face recognition," *Proc. of International Conference on Pattern Recognition*, pp. 306-309, vol. 4, 2002.

[15] P. Yang, S. Shan, W. Gao, S. Li, and D. Zhang, "Face recognition using Ada-boosted Gabor features," *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 356- 361, 2004.

[16] S. Shan, P. Yang, X. Chen, and W. Gao, "Adaboost Gabor Fisher classifier for face recognition," *AMFG 2005*, *LNCS 3723*, pp. 278-291, 2005.

[17] B. Gokberk, M. Irfanoglu, L. Akarun, and E. Alpaydm, "Optimal Gabor kernel location selection for face recognition," *Proc. of International Conference on Image Processing*, vol. 1, pp. 677-680, 2003.

[18] B. Kepenekci, F. Boray Tek, and G. Bozdagi Akar, "Occluded face recognition based on Gabor wavelets," *Proc. of International Conference on Image Processing*, vol. 1, pp. 293-296, 2002.

[19] N. Kruger, "An algorithm for the learning of weights in discrimination functions using a priori constraints," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 764-768, 1997.

[20] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.

[21] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, 1991.

[22] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103-108, 1990.

[23] A. Martinez and A. Kak, "PCA versus LDA," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228-233, 2001.

[24] The ORL Database (http://www.uk.research.att.com/facedatabase.html).

[25] The Normalized Yale Face Database (http://vismod.media.mit.edu/vismod/classes/mas622-00/datasets/).

[26] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, 2001.

[27] K. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27,

no. 5, pp. 1-15, May 2005.

[28] S. Shan, W. Gao, Y. Chang, Bo Cao, and P. Yang, "Review the strength of Gabor features for face recognition from the angle of its robustness to mis-alignment," *Proc. of International Conference on Pattern Recognition*, vol. 1, pp. 338-341, 2004.

[29] P. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image and Vision Computing*, vol. 16, no. 5, pp. 295-306, 1998.

[30] P. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, 2000.

[31] V. Blanz, P. Grother, P. Phillips, and T. Vetter, "Face recognition based on frontal views generated from non-frontal Images," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 454-461, 2005.

[32] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063-1074, 2003.

[33] T. Vetter and T. Poggio, "Linear object classes and image synthesis from a single example image," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 733-742, 1997.

CHAPTER 3 ADAPTIVE REGION-BASED IMAGE ENHANCEMENT METHOD FOR ROBUST FACE RECOGNITION UNDER VARIABLE ILLUMINATION CONDITIONS[2]

3.1 Introduction

Due to the difficulties in controlling the lighting conditions in practical applications, the resulting variability in image illumination is one of the most challenging problems in face recognition [1]-[3]. The performance of existing techniques is heavily subject to the variations in the lighting conditions. Over the last decade, some approaches that attempt to tackle the problem of face recognition under varying illuminations have emerged [4]-[15].

The first class of approaches that deal with illumination variations seek to utilize features that are invariant to illumination variations. Edge maps [4], image intensity derivatives [5] and Gabor-like filters [6] have been proposed. However, Adini's empirical studies [7] showed that none of these representations is sufficient to overcome image variations due to large changes in the direction of lighting. In [8], the quotient image was regarded as the illumination invariant signature image which can be used for face recognition under varying illumination 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.

The second class of approaches use face variation modeling. They assume that illumination variations are mainly due to the 3D shape of human faces under lighting in different directions. Recently, some researchers have attempted to construct a generative 3D face model that can be used to render face images with different poses and under varying illumination conditions [9]-[12]. A generative model called illumination cone was presented

[2] A version of this chapter has been submitted for publication, Shan Du and Rabab Ward, "Adaptive Region-based Image Enhancement Method for Robust Face Recognition under Variable Illumination Conditions."

in [9][10]. The main idea of this method is that the set of face images with fixed pose but under different illumination conditions can be represented using an illumination convex cone which can be constructed from a number of images acquired under variable lighting conditions. The illumination cone can be approximated in a low-dimensional linear subspace. In [11], it was shown 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. In [12], a method called spherical harmonic-based representations was proposed. It requires the knowledge of either the light source or a large volume of training data, which is not practical for most real world scenarios. One of the drawbacks of the model-based approaches is that they need a number of images of a subject under varying lighting conditions or 3D shape information during the training phase. This drawback limits its applications in practical face recognition systems.

The third class of approaches use face image preprocessing techniques to normalize the images under different illumination conditions. For example, histogram equalization (HE) and Gamma intensity correction (GIC) are widely used for illumination normalization [13]. However, uneven illumination variation is still difficult to cope with by using these global processing techniques. Recently, region-based histogram equalization (RHE) [13] and block-based histogram equalization (BHE) [14] have been proposed to deal with uneven illumination variations. Although the recognition rates can be improved compared with HE, their performance is still not satisfactory. In [13], 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. In [15], by combining symmetric shape-from-shading (SSFS) method and a generic three-dimensional

(3D) model, the performance of face recognition under varying illuminations was improved. However, this method is only efficient for exact frontal face images and it is assumed that all faces share a similar common shape. These assumptions limit their effectiveness in practical applications.

In this chapter, we present a new preprocessing illumination adjustment method, where 3D modeling and thus extra images used for model fitting are not needed. To solve the illumination variation problem, especially the side lighting problem, we propose a novel adaptive region-based image preprocessing scheme that enhances face images and facilitates the illumination invariant face recognition task.

The proposed method first automatically segments an image into differently lit regions according to its different local illumination conditions, then for every region, both the contrast and the edges are enhanced separately so as to alleviate the highlight, shading and shadow effects caused by uneven illuminations.

The region segmentation of the face image is based on an edge map obtained using a new wavelet-based edge extraction method. This is because the intensity of edges of a region reflects the illumination conditions of that region and wavelets have the merit of extracting edges. In order to take advantage of the multi-resolution property of the wavelet transform, we propose multiplying the corresponding detail coefficients of two adjacent resolutions to extract image edges.

Since illumination variations mainly lie in the low-frequency band, after segmenting the image into differently lit regions, histogram equalization (HE) of the low-frequency coefficients is then applied regionally to minimize the variations under different lighting

conditions. This proposed contrast enhancement method is called Adaptive Region-based Histogram Equalization (ARHE).

We also notice that under poor illuminations the high-frequency features become more important in recognition. Therefore, for every differently lit region, its high-frequency coefficients are enlarged to make face images more distinguishable. This proposed region-based edge enhancement procedure is called EdgeE in this chapter.

The experimental results using the combination of ARHE and EdgeE on the representative Yale B database show that the proposed approach significantly improves the recognition performance of face images with illumination variations. Other advantages of our approach include its simplicity and generality. It does not require any modeling and model fitting steps and can be implemented easily. It can also be applied directly to any single image without any lighting assumption or any prior information about the 3D face geometry.

The remainder of this chapter is organized as follows. In Section 3.2, the framework of the proposed method is presented. Section 3.3 describes the proposed region segmentation method. Section 3.4 presents the descriptions of the proposed region-based image enhancement method based on edges (EdgeE) and histogram equalization (ARHE). Section 3.5 illustrates our image enhancement experiments on Yale B face images and shows the face recognition results. Section 3.6 concludes this chapter.

3.2 Framework of the Proposed Method

The first step of our method is segmenting an image into regions with different local illumination conditions. The reason why we do this is as follows: When side lighting effect exists, different regions of an image possess different illumination properties. The regions facing the light source may be well-lit, while the regions not directly facing the light source

may be under-lit. Regions with attached shadows caused by under lighting need to be separated (and to be dealt with separately) from the well-lit regions. Moreover, due to the face geometry, some regions may be blocked from the lighting source and appear dark. Regions with the so-called cast shadows also need to be processed separately. Therefore, the region segmentation step first automatically partitions an image into regions where every region has homogeneous illumination property. Then every region is processed according to its local illumination condition so as to alleviate the highlight, shading and shadow effects caused by uneven illumination. When an image has even lighting, the illumination property of the whole image is homogeneous; the segmentation step only results in one region, which is the whole image. Thus, our method is adaptive to the actual image illumination conditions, and is region-based.

The whole face recognition system is described by the block diagram in Figure 3-1. First, the wavelet transform is employed to decompose an image into approximation coefficients (low-frequency components) and multi-level detail coefficients (high-frequency components). Then an edge map of the input image is generated using a new wavelet-based edge extraction method proposed in this chapter. The edge map is used to segment the image into regions with different illumination conditions. The reason why we use edge map to do segmentation is explained in Section 3.3.3. Based on the edge map, the input image is segmented into regions with homogeneous illumination conditions. For each region, contrast enhancement using the ARHE method on the approximation coefficients corresponding to that region and edge enhancement using the EdgeE method on the detail coefficients also corresponding to that region are carried out separately. At the end, an enhanced image is reconstructed using inverse the discrete wavelet transform (IDWT) with the modified approximation coefficients

and the modified detail coefficients. The reconstructed enhanced image is then used in the face recognition system for final classification. The Euclidean distance nearest-neighbor classifier is used to find the best match image in the database with the input image.

More details of edge generation, region segmentation, ARHE and EdgeE will be discussed in the following sections.

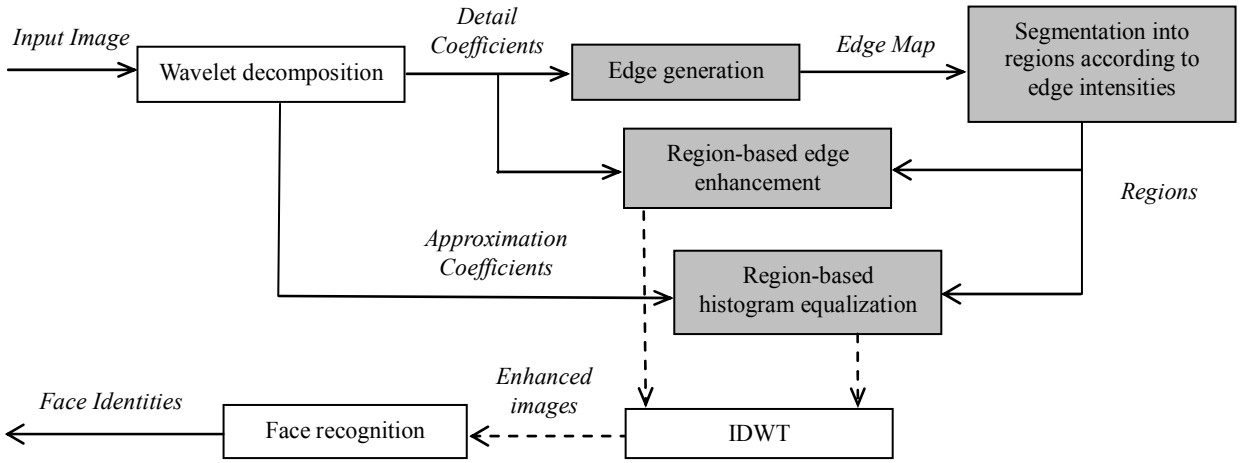


Figure 3-1. Block diagram of the proposed adaptive region-based image enhancement method for face recognition (the grey blocks are the major contributions).

3.3 Region Segmentation

Region segmentation is based on an edge map of the face image. The edge map is obtained using a new wavelet-based edge extraction method proposed below. The reason behind using an edge map in segmentation is because we have observed that uneven illumination conditions affect the intensities of edges of different local regions of an image differently. Lack of light or over lighting weakens the edges in the edge map (as shown in Figure 3-5 (d) and (f) in Section 3.3.2). Therefore, we examine the edge map. If some edges are weak, then the illumination in this region could be either over-lit or under-lit. We differentiate these regions from the normally lit regions and apply different contrast

enhancement as well as edge enhancement on each region separately. The edge maps are generated using wavelet decomposition.

3.3.1 Wavelet Decomposition

Edges involve the high-frequency components of images. In practice, calculating the high-frequency components of an image is achieved by convolving a kernel with the image. The effectiveness of edge extraction is decided by the kernel size. Usually the kernel size is chosen by the trial and error method. Once the kernel size is fixed, the resolution is fixed.

But often, edges occur at different resolutions; both strong edges and weak edges exist in the same image. It is difficult to choose a kernel size to extract all the edges. So it is appropriate to extract edges at different scales or resolutions.

Wavelet decomposition [16] represents a given signal as a set of basis functions at different scales. Since wavelets are short-time oscillatory functions having finite support length (limited duration both in time and frequency), they are localized in both the time (spatial) and the frequency domains. The joint spatial-frequency resolution obtained by the wavelet transform makes it a good candidate for the extraction of details as well as approximations of images.

In the two-band multi-resolution wavelet transform, a signal can be expressed by wavelet and scaling basis functions at different scales, in a hierarchical manner, that is

$$f(x) = \sum_{k} a_{0,k}\phi_{0,k}(x) + \sum_{j}\sum_{k} d_{j,k}\psi_{j,k}(x) \qquad (3\text{-}1)$$

where $\phi_{j,k}$ are scaling functions at scale j and $\psi_{j,k}$ are wavelet functions at scale j. $a_{j,k}$, $d_{j,k}$ are scaling coefficients (approximation) and wavelet coefficients (detail).

For the 2D discrete wavelet transform (2D DWT), the approximation coefficients (low-frequency components) and detail coefficients (high-frequency components) can be easily computed using a 2D filter bank consisting of low-pass and high-pass filters. After one level of 2D decomposition, an image is divided into four sub-bands: LL (Low-Low), which is generated by the approximation coefficients; LH (Low-High), HL (High-Low), and HH (High-High), which are generated by the detail coefficients, as shown in Figure 3-2.

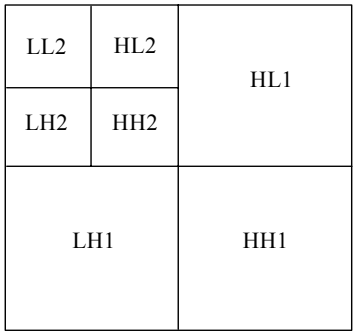


Figure 3-2. Multi-resolution structure of wavelet decomposition of an image.

After applying the wavelet transform, the given image is decomposed into several frequency components in multi-resolution. Using different wavelet filter sets and/or different number of transform-levels will result in different decomposition results. Since selecting wavelets is not our focus in this chapter, we choose the 2-level db1 wavelets in our experiments. However, any wavelet-filters can be used in our proposed method.

For the edge extraction purpose, in this chapter, we use the redundant wavelet transform replacing the normal wavelet transform. The decomposition procedure for a redundant wavelet transform is different from the normal one in that the scaling of the wavelet is not achieved by sub-sampling the image in each step, but rather by an up-sampling of the filters. The four wavelet sub-bands at scale j are of the same size as the original image, and all filters

used at scale j are up-sampled by a factor of 2^j (padding $2^j - 1$ zeros) compared with those at scale zero. Figure 3-3 shows the 2-level redundant wavelet decomposition of a face image.

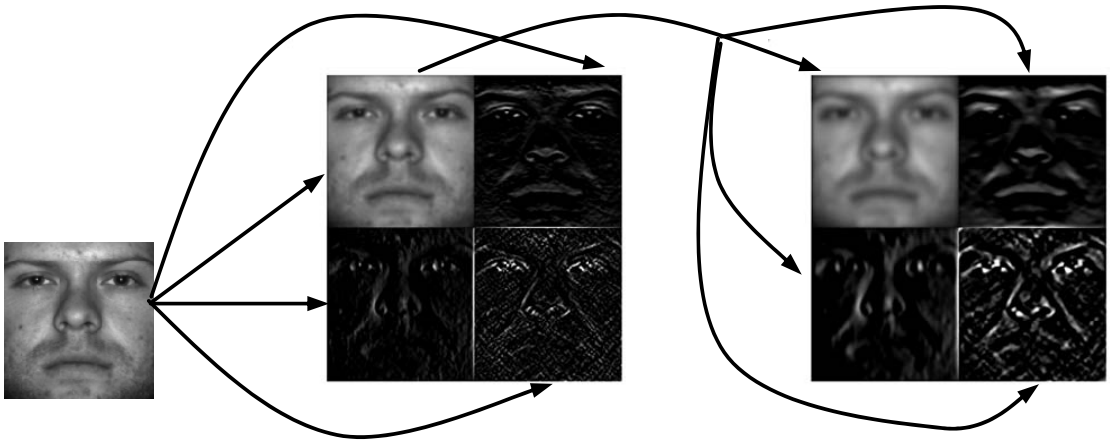


Figure 3-3. 2-level redundant wavelet decomposition of a face image.

3.3.2 Edge Map Generation

The edges of an image are full of high-frequency information about the image that will scatter into several scales or resolutions. In order to take advantage of the multi-resolution property of the wavelet transform, we propose multiplying each pair of corresponding sub-bands at adjacent resolutions to enhance image edges and suppress noise. This is based on the fact that edge structures are present at each scale while noise decreases rapidly along the scales.

Multiplication of the corresponding detail coefficients of two adjacent levels is shown in Figure 3-4. We compute LH1×LH2, HL1×HL2, and HH1×HH2. The edge map E is obtained by

$$E = \sqrt{(LH1 \times LH2) + (HL1 \times HL2) + (HH1 \times HH2)} \qquad (3\text{-}2)$$

Figure 3-5 shows three examples of edge maps of three differently illuminated face images.

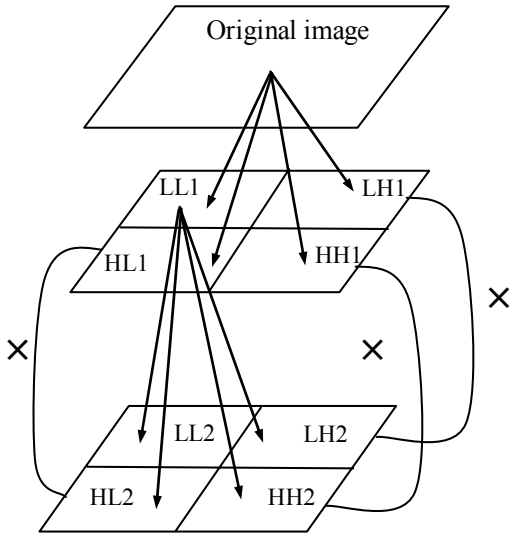


Figure 3-4. Edge generation by multiplying corresponding detail coefficients at two adjacent decomposition levels.

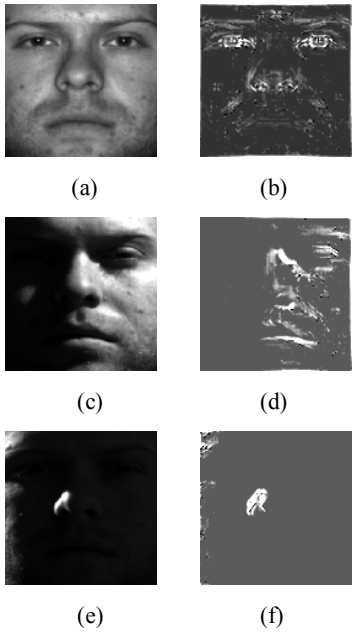


Figure 3-5. Edge maps of three differently illuminated face images.

(a), (c) and (e): original images; (b), (d) and (f): edge maps

3.3.3 Region Segmentation

To segment an image, we use our observation that under uneven illumination conditions, the edges at different regions have different intensity strengths. Lack of light or over lighting weakens the edges. Therefore, we examine the edge map. If a region's edge intensity is

normal, it means this region is lit normally; but if some edges in the edge map are weak, then the illumination in this region may be over-lit or under-lit. We differentiate these regions from the well-lit regions. Then we apply edge enhancement as well as contrast enhancement on each region separately. Here we should note that if a region's edge intensity is weak, besides the lighting problem, this may be caused by a natural solid color, e.g., the cheeks. Thus, we need to differentiate these two cases.

Our algorithm for segmenting the face regions is shown in Figure 3-6.

1. *Partition an edge map into blocks, e.g., 16×16.*
2. *For each block, compute the average edge amplitude:*
   a. *If the edges in the block are weak (e.g., less than the global average of the whole image), merge it with adjacent weak blocks using the eight nearest neighbours method and then mark it with 0;*
   b. *Otherwise, mark it with 1.*
3. *For a region marked with 0:*
   a. *If its grey-scale is similar to any of the regions marked with 1, then, mark them back to 1. This step removes the regions that have weak edges not due to light, but due to a natural solid color, e.g., the cheeks;*
   b. *Otherwise, this region is affected by illumination changes.*
4. *For the regions marked with 1, consider them as one whole region, the normal region.*
5. *End.*

Figure 3-6. Region segmentation algorithm.

In Figure 3-7, the segmented regions of three differently illuminated images are shown using blue lines. Here, we should note that the number of regions may be greater than or equal to 1. For an evenly illuminated image, the number of regions is only 1. For an unevenly illuminated image, the number of regions is greater than 1.

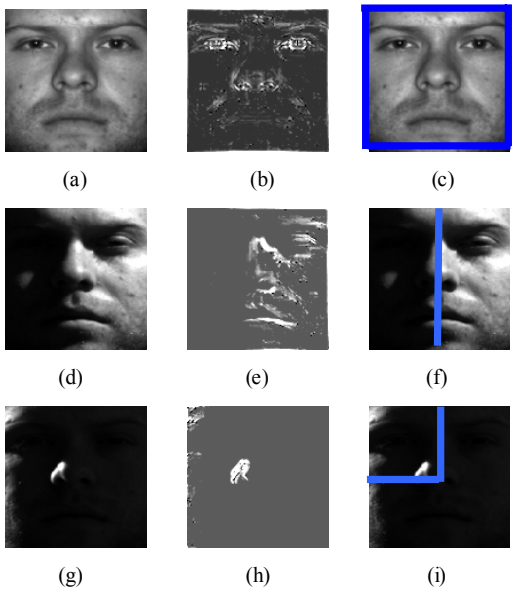


Figure 3-7. Segmented regions of three differently illuminated face images (separated by blue lines).

(a), (d) and (g): original images; (b), (e) and (h): edge maps; (c), (f), and (i): segmented regions obtained based on edge maps

3.4 Region-based Image Enhancement for Face Recognition

Our proposed region based image enhancement method enhances both the contrast and the edges for each region separately so as to alleviate the side lighting effect caused by the uneven illumination conditions. Contrast enhancement is done by the proposed adaptive region-based histogram equalization (ARHE) carried on the approximation coefficients (low-frequency components) since illumination variations mainly lie in the low-frequency band. As well, we propose enhancing the edges (EdgeE) by enlarging the detail coefficients (high-frequency components) at different scales since with poor illuminations the high-frequency features become more important in recognition.

3.4.1 Adaptive Region-based Contrast Enhancement (ARHE)

In the existing literature, contrast enhancement is achieved by applying histogram equalization (HE) on a whole gray-scale image so as to redistribute gray levels uniformly.

However, after processing by HE, the lighting condition of an image with uneven illumination sometimes becomes worse, i.e., more uneven. This is because HE is a global transform applied over the whole image area and therefore, may be less effective when side lighting exists.

Unlike all other papers using HE, in this chapter, a region-based histogram equalization is used to enhance the contrast. Histogram equalization of the approximation coefficients is applied separately on each segmented region and not on the whole image to avoid over lighting the already normally lit regions or under lighting the dark regions. The regions of an image are produced by the edge map generation and region segmentation algorithms described in Section 3.3.

If there is no sub-region in one image, i.e., the segmentation method does not result in more than one region, which means the image is evenly illuminated, then, of course, the image is processed holistically.

The reason why we do histogram equalization on the approximation coefficients is because illumination variations mainly lie in the low-frequency band. Thus, histogram equalizing the low-frequency coefficients can minimize the variations arising from different lighting conditions. Figure 3-8 shows the regionally contrast enhanced approximation coefficients on differently illuminated images.

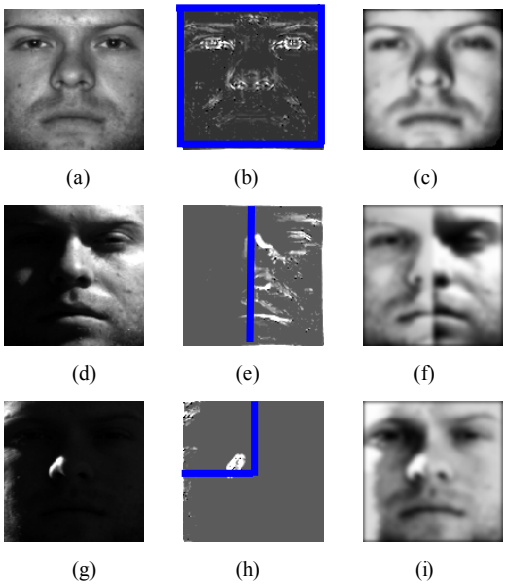


Figure 3-8. Regionally contrast enhanced approximation coefficients.

(a), (d) and (g): original images; (b), (e) and (h): segmented regions obtained using edge maps; (c), (f), and (i): enhanced approximation coefficients by ARHE

3.4.2 Adaptive Region-based Edge Enhancement (EdgeE)

We notice that under poor illuminations, the high-frequency features become more important in recognition. Therefore, the detail coefficients (edges) need to be enhanced to make face images more distinguishable. However, to the best of our knowledge, so far, no similar observation and solution have been reported in literature.

The proposed edge enhancement attempts to emphasize the fine details in the original image. The perceptibility of edges and small features can be improved by raising the amplitude of high-frequency components in the image.

The edges of some regions of the original image may be very weak due to under lighting or over lighting conditions (the amplitude of the detail coefficients may be around zero); thus it is hard to enlarge the edges so that they have similar magnitudes as the ones obtained from well-lit regions. Here, we do not enhance the original detail coefficients, rather we enhance the detail coefficients obtained using the regionally re-lit (i.e., illumination adjusted) images

as shown in Figure 3-9. The re-lit images are obtained by applying HE on the gray levels of each region separately. Figure 3-10 shows the detail coefficients obtained from the original badly-illuminated images. Figure 3-11 shows the detail coefficients obtained from the regionally re-lit images.

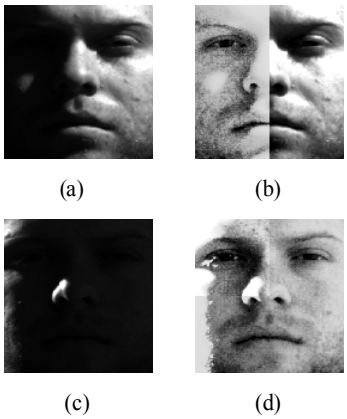


Figure 3-9. Regionally re-lit images.
(a), (c): original images; (b), (d): regionally re-lit images

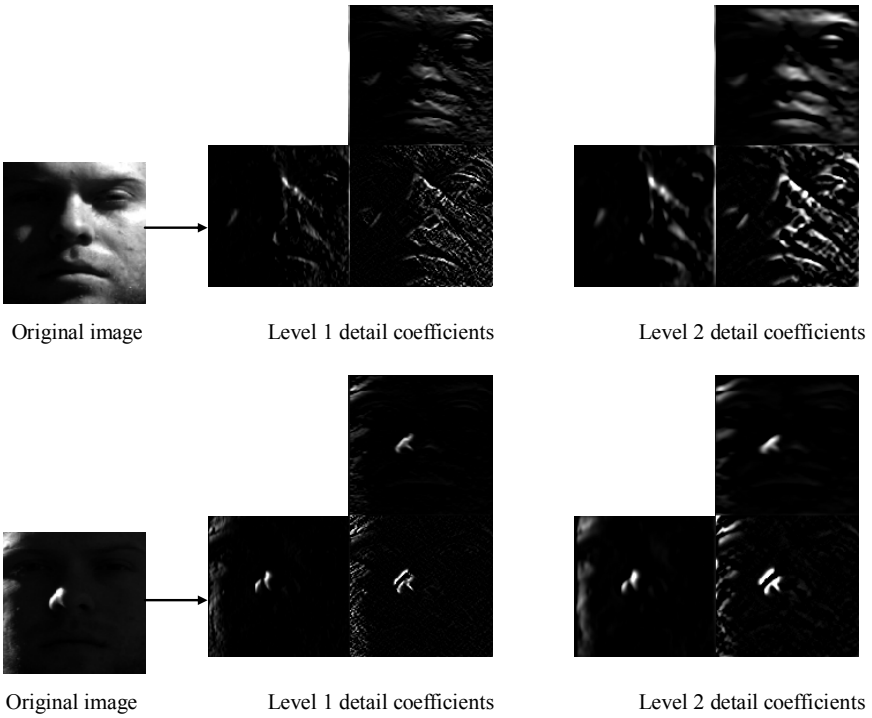


Figure 3-10. Detail coefficients obtained using the original images.

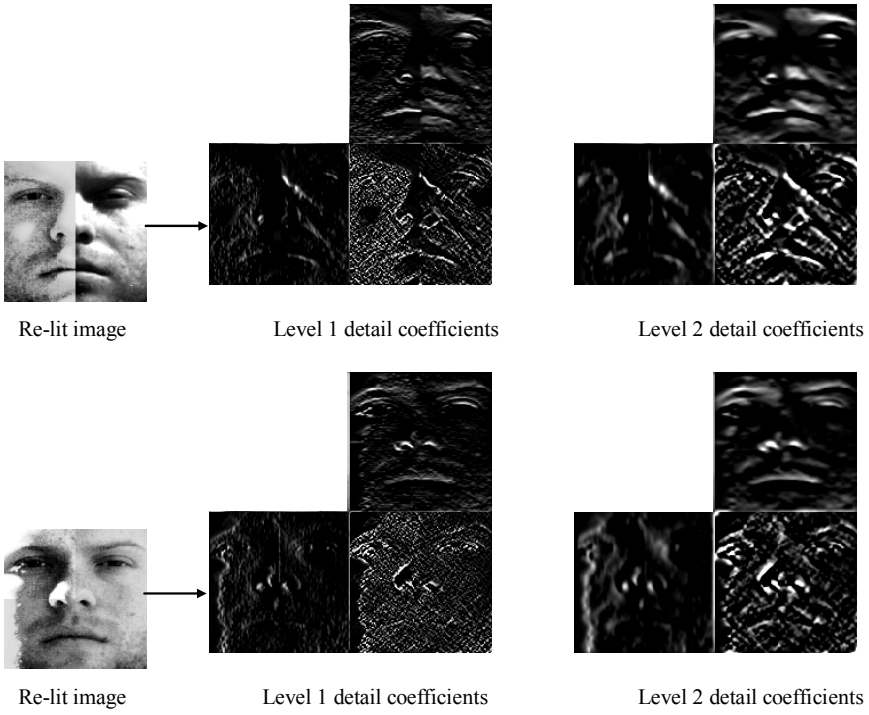


Figure 3-11. Detail coefficients obtained using the re-lit images.

After the above adjustment, the newly generated edges of the different regions may still have different strengths. To make the edges of the different regions of even strengths, we introduce an adaptive edge enhancement method, EdgeE, based on the region segmentation results. We propose using different enlargement factors for the different regions based on their original edge amplitudes.

Let m_i denote the average edge amplitude of the i^{th} region of the new edge map E_n obtained using the re-lit images. The enhanced new edge map EE_k for each region k can be calculated by

$$EE_k(x,y) = \frac{E_{nk}(x,y)}{m_k} \times \max\{m_1, m_{2,}..., m_n\} \times \alpha \qquad (3\text{-}3)$$
$$(x,y) \in region \quad k$$

where E_{nk} is the edge map of the re-lit image at region k, α is a scaling factor, it is greater

than 1.

Thus, for the detail coefficients of each region *k*, the enlargement factor f_k of this region is

$$f_k = \frac{EE_k(x,y)}{E_{nk}(x,y)} \tag{3-4}$$

3.4.3 Face Recognition

After the above preprocessing stage that consists of region segmentation, contrast enhancement and edge enhancement of every region separately, the inverse discrete wavelet transform (IDWT) is applied on the modified approximation coefficients and the modified detail coefficients to obtain a reconstructed enhanced image. The reconstructed enhanced image is then used in a face recognition system for final classification. In the experiments below, the face recognition employs the Euclidean distance nearest-neighbor classifier to find the image in the database with the best match.

3.5 Experimental Results

3.5.1 Yale Face Database B

To evaluate the performance of the proposed image preprocessing method, we test it on the Yale Face Database B [10], which was built by the Center for Computational Vision and Control at Yale University. Yale B is a representative face image database used widely to evaluate face recognition techniques. It contains 5760 single light source images of 10 subjects. Each subject has 9 poses and each pose has 64 different illumination conditions. Since this chapter mainly deals with the illumination problem, the frontal pose images captured under 64 different lighting conditions are chosen. Example images of one person in frontal pose are shown in Figure 3-12. The images are divided into five subsets according to

the light-source directions (azimuth and elevation): Subset 1 (angle < 12 degrees from optical axis), Subset 2 (20 < angle < 25 degrees), Subset 3 (35 < angle < 50 degrees), Subset 4 (60 < angle < 77 degrees), and Subset 5 (others).

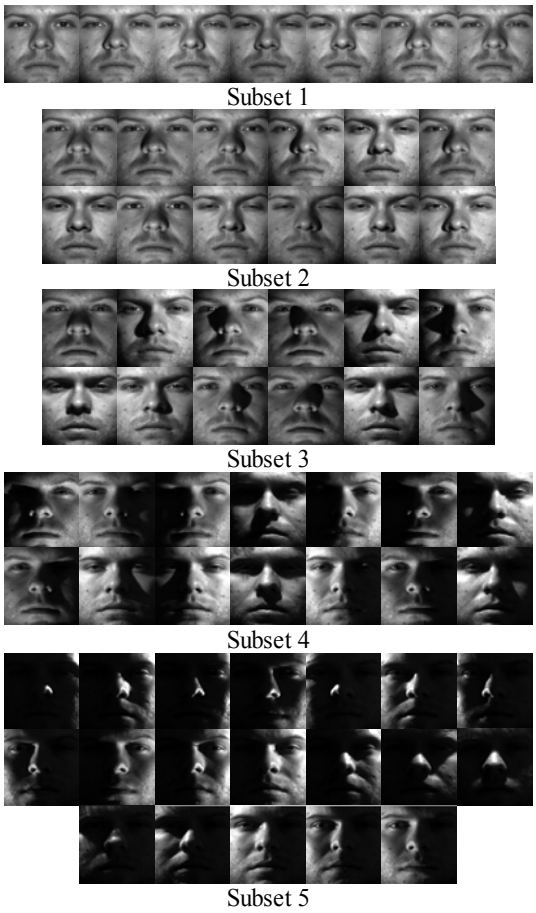


Figure 3-12. 64 illumination conditions for one person.

3.5.2 Parameters Setting

In the experiments, there are three parameters to be set.

3.5.2.1 Wavelet decomposition level

In our method, wavelet transform is used to decompose an image into approximation coefficients and detail coefficients, to generate edge maps, and the inverse wavelet transform is used to reconstruct the enhanced image. Although, the selection of wavelet basis function and decomposition level is not the focus of this chapter, we do test the influence of the

decomposition level to the final recognition performance. We test 1 to 4 decomposition levels, 2-level decomposition is better than 1-level decomposition in terms of recognition performance. More decomposition levels, e.g., 3 or 4 levels do not generate better results any more but take more time to do the decomposition and reconstruction, so we choose 2-level decomposition.

3.5.2.2 Segmentation block size

In region segmentation, we first segment an edge map into blocks. The block size is another consideration. There is a tradeoff in the selection. If the block size is too large, we may miss some small specific regions; if the block size is too small, the segmentation will take longer time. This parameter only depends on the scale of an image, and is very easy to set. In other words, one parameter can be used for all face images approximately of the same size. Because the lighting changes gradually, the homogeneous regions normally possess a certain area of an image; normally 1/2 or 1/4. In our experiments, a block size of 1/16 of the image size is small enough to catch all small independent regions. For example, for a 64×64 image, the block size of 16×16 is appropriate.

3.5.2.3 Detail coefficients enlargement factor α

As aforementioned, high-frequency coefficients should be enlarged to make face images more distinguishable. Then how large is the enlargement factor α? In the experiments, we test α value from 2 to 100. The reconstructed enhanced images using different α values are used to compute the ratio of the between-class variance to the within-class variance. When the ratio is large, it means the corresponding recognition rate is high. The α value corresponding to the highest between-class variance to within-class variance ratio is chosen as the optimal choice.

Let $h \times w$ denotes the size of the reconstructed enhanced image $I(\vec{x}_{h,w})$; $\sigma^2(\vec{x}_{h,w})$ be the ratio of the between-class variance to the within-class variance of the reconstructed enhanced image. To calculate $\sigma^2(\vec{x}_{h,w})$, we use a set of K example images of c classes (different persons) $\{X_1, X_2, ..., X_c\}$, and each class has K_i images. The gray-scale values of the K example images are $I^k(\vec{x}_{h,w})$, $k = 1 \to K$. The gray-scale values of the K_i example images of class X_i are $I^{k_i}(\vec{x}_{h,w})$, $k_i = 1 \to K_i$.

The between-class variance of $I(\vec{x}_{h,w})$ is defined as

$$\sigma_B{}^2(\vec{x}_{h,w}) = \frac{\sum_{i=1}^{c} K_i [\mu_i(\vec{x}_{h,w}) - \mu(\vec{x}_{h,w})]^2}{K} \tag{3-5}$$

where $\vec{x}_{h,w}$ is the pixel position; $\mu_i(\vec{x}_{h,w})$ is the mean of the K_i reconstructed enhanced image $I^{k_i}(\vec{x}_{h,w})$, $k_i = 1 \to K_i$ of class X_i; $\mu(\vec{x}_{h,w})$ is the mean of the K reconstructed enhanced image $I^k(\vec{x}_{h,w})$, $k = 1 \to K$; K is the number of example images and K_i is the number of samples in class X_i; c is the number of classes; $K = \sum_{i=1}^{c} K_i$. $\mu_i(\vec{x}_{h,w})$ and $\mu(\vec{x}_{h,w})$ are expressed by

$$\mu_i(\vec{x}_{h,w}) = \frac{\sum_{k_i=1}^{K_i} I^{k_i}(\vec{x}_{h,w})}{K_i} \tag{3-6}$$

$$\mu(\vec{x}_{h,w}) = \frac{\sum_{k=1}^{K} I^{k}(\vec{x}_{h,w})}{K} = \frac{\sum_{i=1}^{c} \sum_{k_i=1}^{K_i} I^{k_i}(\vec{x}_{h,w})}{\sum_{i=1}^{c} K_i} \tag{3-7}$$

The within-class variance is defined as

$$\sigma_W{}^2(\vec{x}_{h,w}) = \frac{\sum_{i=1}^{c} \sum_{k_i=1(I^{k_i}(\vec{x}_{h,w})\in X_i)}^{K_i} [I^{k_i}(\vec{x}_{h,w}) - \mu_i(\vec{x}_{h,w})]^2}{K} \tag{3-8}$$

where $\vec{x}_{h,w}$ is the pixel position; $\mu_i(\vec{x}_{h,w})$ is the mean of the K_i reconstructed enhanced image $I^{k_i}(\vec{x}_{h,w})$, $k_i = 1 \to K_i$ of class X_i; K_i is the number of samples in class X_i; c is the number of classes.

The between-class variance to within-class variance ratio $\sigma^2(\vec{x}_{h,w})$ of the reconstructed enhanced image $I(\vec{x}_{h,w})$ is then given by

$$c = \frac{\sigma_B{}^2(\vec{x}_{h,w})}{\sigma_W{}^2(\vec{x}_{h,w})} \tag{3-9}$$

To compute the ratio of the between-class variance to the within-class variance, we use all the reconstructed enhanced images of the Yale B database as the example images. There are 10 classes (person) and each class has 64 images. Then,

$$\alpha_{optimal} = Arg \max_j(c_j), j = 2...100 \tag{3-10}$$

In Figure 3-13, we show the relationship of detail coefficients enlargement factor α with the ratio of between-class variance to within-class variance. From the chart, we find that $\alpha = 50$ corresponds to the highest between-class variance to within-class variance ratio. Thus, we can get the best recognition result with $\alpha = 50$. From our recognition experiments, we do get the best recognition rate with $\alpha = 50$. They are consistent.

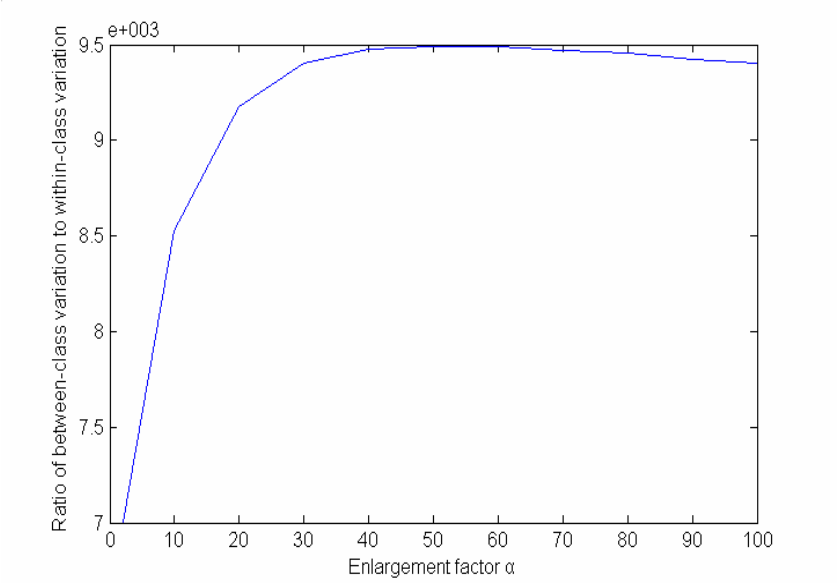


Figure 3-13. The relationship of detail coefficients enlargement factor α with the ratio of between-class variance to within-class variance.

3.5.3 Image Enhancement Results

The enhancement results after applying the above preprocessing scheme to three differently illuminated images are shown in Figure 3-14, Figure 3-15 and Figure 3-16. The sub-figures (c), (d) and (e) in each figure show the images enhanced by HE, ARHE, and the combination of ARHE and EdgeE respectively.

The histogram equalized image (c) in Figure 3-15 shows that when uneven lighting exists, the global histogram equalization method makes the already normally lit regions over lit while the dark regions are still dark. Our region-based method processes each region separately so that all regions become pronounced enough for automatic face recognition.

We should note that although some boundaries between the regions show up in the enhanced images by the region-based method, they do not influence the recognition performance. In face recognition applications, we are more concerned with recognition

performance rather than visual image quality. The experimental results in Section 3.5.4 will show the advantages of using region-based method.

Figure 3-17 shows the enhanced images of the original Yale B images of one person under 64 illumination conditions and Figure 3-18 shows the enhance images of the original images of 10 persons.

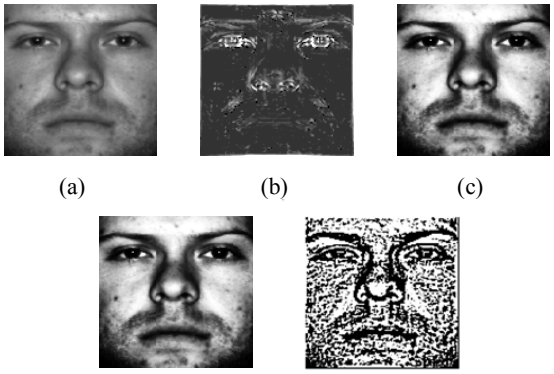


Figure 3-14. (a) **evenly** illuminated face image; (b) its edge map; (c) enhanced image by HE; (d) enhanced images by ARHE, and (e) enhanced images by ARHE+EdgeE.

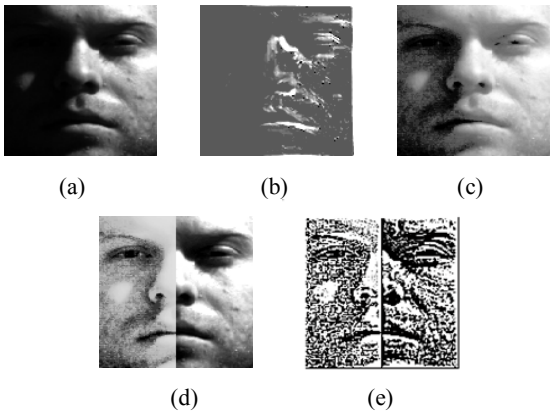


Figure 3-15. (a) **unevenly** illuminated face image; (b) its edge map; (c) enhanced image by HE; (d) enhanced images by ARHE, and (e) enhanced images by ARHE+EdgeE.

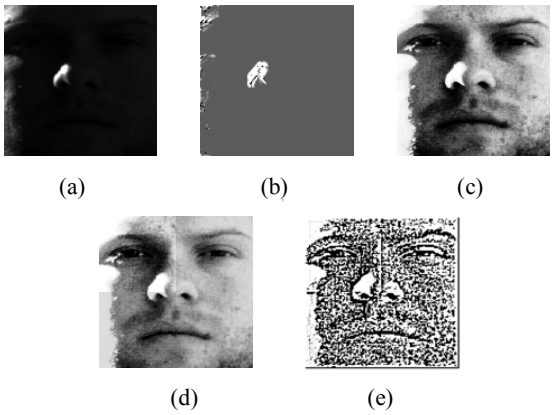


Figure 3-16. (a) **badly** illuminated face image; (b) its edge map; (c) enhanced image by HE; (d) enhanced images by ARHE, and (e) enhanced images by ARHE+EdgeE.

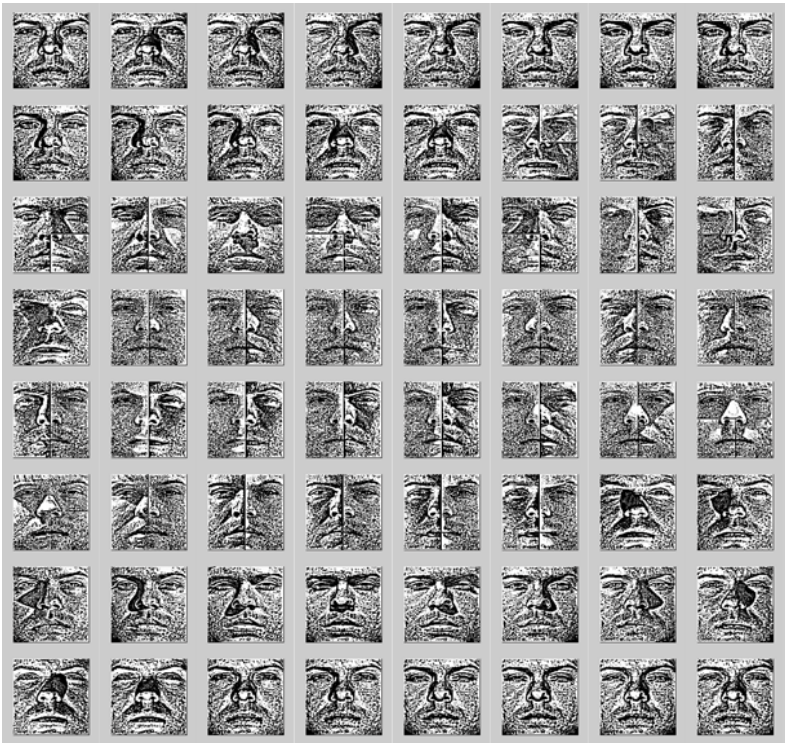


Figure 3-17. ARHE+EdgeE enhanced images of the original Yale B images of one person under 64 illumination conditions.

Figure 3-18. ARHE+EdgeE enhanced images of the original images of 10 persons.

3.5.4 Recognition Results

In our experiments, Subset 1 (7 images for each person) is chosen as the gallery and each of the images in the remaining 4 subsets is matched to each of the images in the gallery so as to find the best match using the Euclidean distance nearest-neighbor classifier.

The same recognition system is applied on the same set of images but after preprocessing them by different algorithms. Recognition performances using the different preprocessing algorithms are compared in this section. The abbreviations for the preprocessing algorithms are:

Raw image: no preprocessing, i.e., recognition system is applied on the original images

HE: conventional global Histogram Equalization

RHE [13]: uniform Regional Histogram Equalization (divide images into 4 equal sub-regions, and histogram equalize the gray-scale values of every sub-region)

ARHE: Adaptive Region-based Histogram Equalization

ARHE+EdgeE: combination of ARHE and the adaptive region-based Edge Enhancement

The recognition rates are illustrated in Table 3-1 and Figure 3-19. It is shown from Table 3-1 that our proposed method, ARHE+EdgeE, outperforms the other methods in every single subset. Even without EdgeE, our proposed ARHE outperforms RHE and HE significantly. It means the adaptive region-based contrast enhancement method is better than the conventional methods. Edge enhancement added to face recognition can further improve the recognition rate significantly.

Table 3-1. Recognition rate comparisons of different preprocessing methods on Yale face database B (Subset 1 is used as the gallery).

Methods	Subset 2	Subset 3	Subset 4	Subset 5	Average
Raw image	95.83	76.67	46.67	25.24	55.65
HE	100	97.50	75	60	79.47
RHE [13]	100	100	84.17	65.71	84.03
ARHE	100	100	90	80	90.53
ARHE+EdgeE	**100**	**100**	**99.16**	**96.66**	**98.59**

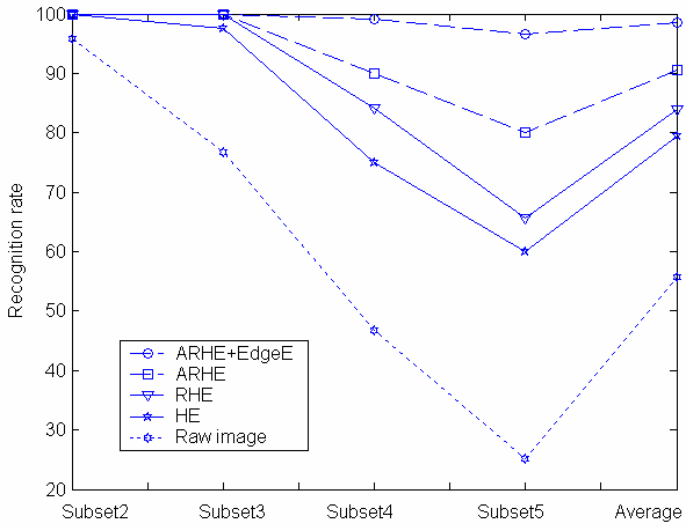


Figure 3-19. Recognition rate comparisons of different preprocessing methods.

3.6 Conclusions

In this chapter, in order to solve the varying illumination problem, especially the side lighting effect problem in face recognition, we propose a novel adaptive region-based image preprocessing scheme that enhances face images and facilitates the illumination invariant face recognition task. The proposed method first segments an image into different regions according to its different local illumination conditions, then both the contrast and the edges are enhanced regionally so as to alleviate the side lighting effect. Since illumination variations mainly lie in the low-frequency band, the proposed contrast enhancement scheme uses adaptive region-based histogram equalization (ARHE) of the low-frequency coefficients

to minimize variations under different lighting conditions. By observing that under poor illuminations the high-frequency features are more important in recognition, we enlarge the high-frequency coefficients (EdgeE) to make face images more distinguishable.

Compared with existing methods, our new method is more suitable for dealing with uneven illuminations in face images. Experimental results show that the proposed method improves the performance significantly when the face images have illumination variations. The other advantages of our method include the following: it does not require any 3D modeling and model fitting steps and can be implemented easily; it can be applied directly to any single image without any lighting assumption, any prior information on 3D face geometry.

The major contributions of this chapter are summarized as follows. (1) It is the first work that uses edge enhancement method in face recognition, (2) It is the first work that uses region-based image enhancement method that adapts to the actual illumination conditions of images, (3) Image edge map generation and region differentiation/segmentation algorithms are proposed. The region segmentation is based on its edges rather than its intensity values, and (4) Both even and uneven lighting conditions are considered and processed accordingly and automatically.

From the experimental results, we conclude that: (1) Adaptive region-based image enhancement substantially outperforms the existing illumination normalization methods, (2) Edge enhancement (EdgeE) plays an important role in face recognition, (3) Combining together the edge enhancement and contrast enhancement (ARHE) offers more benefit.

3.7 References

[1] R. Chellappa, C. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proc. of the IEEE*, vol. 83, no. 5, pp. 705-740, 1995.

[2] A. Samal and P. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Pattern Recognition*, vol. 25, pp. 65-77, 1992.

[3] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, 2003.

[4] V. Govindaraju, D. Sher, R. Srihari, and S. Srihari, "Locating human faces in newspaper photographs," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 549-554, 1989.

[5] S. Edelman, D. Reisfeld, and Y. Yeshurun, "A system for face recognition that learns from examples," *Proc. of European Conference on Computer Vision*, pp. 787-791, 1992.

[6] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency and orientation, optimized by two dimensional cortical filters," *Journal of the Optical Society of America*, vol. 2, pp. 1160-1169, 1985.

[7] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 721-732, 1997.

[8] A. Shashua and T. Riklin-Raviv, "The quotient image: class-based re-rendering and recognition with varying illuminations," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 129-139, 2001.

[9] P. Belhumeur and D. Kriegman, "What is the set of images of an object under all possible lighting conditions?" *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 270-277, 1996.

[10] A. Georghiades, P. Belhumeur and D. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, 2001.

[11] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218-233, 2003.

[12] L. Zhang and D. Samaras, "Face recognition under variable lighting using harmonic image exemplars," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 19-25, 2003.

[13] S. Shan, W. Gao, B. Cao, and D. Zhao, "Illumination normalization for robust face recognition against varying lighting conditions," *Proc. of IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 157-164, 2003.

[14] X. Xie and K. Lam, “Face recognition under varying illumination based on a 2D face shape model,” *Pattern Recognition*, vol. 38, no. 2, pp. 221-230, 2005.

[15] W. Zhao and R. Chellappa, “Illumination-insensitive face recognition using symmetric shape-from-shading,” *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 286-293, 2000.

[16] Strang and Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1997.

CHAPTER 4 EYE DETECTION IN GRAY-SCALE FACE IMAGES UNDER VARIOUS ILLUMINATION CONDITIONS[3]

4.1 Introduction

Automated detection and segmentation of a face [1][2] has become one of the most important research topics in computer vision and pattern recognition. The motivation behind developing such a system is the great number of its applications. Facial features, such as eyes, eyebrows, nose and mouth, as well as their spatial relationship, are important for facial interpretation tasks including face recognition, facial expression analysis, face tracking, face animation, human machine interaction, video indexing, and model-based video coding.

Due to factors such as illumination, head pose, expression, and scale, the facial features vary greatly in their appearance. The illumination is a particularly difficult obstacle affecting automated detection of facial features. Unfortunately, there exists little work on automated detection under variable illumination conditions, especially for extremely badly illuminated images such as shown in Figure 4-1.

Figure 4-1. Images with extremely bad illumination conditions.

In this chapter, a novel eye detection method for gray-scale face images under various illumination conditions is presented. Localization of eyes is a necessary step for many face recognition systems. Before two face images can be compared, they should be aligned in

[3] A version of this chapter has been submitted for publication, Shan Du and Rabab Ward, "Eye Detection in Gray-scale Face Images under Various Illumination Conditions."

orientation and normalized in scale. Since both the locations of the two eyes and the interocular distance are relatively constant for most people, the eyes are often used for face image normalization. Eye localization also facilitates the detection of other facial features. In our proposed method, the illumination of an image is first regionally adjusted. The re-lighting of the regions makes the feature details more pronounced. The regions are obtained by segmenting the image into different sections of different illumination strengths. The number of regions may be greater than 1 when the illumination is uneven or equal to 1 when the illumination is even. This segmentation is based on an edge map E obtained from the original image I using multi-resolution wavelet transform. Then the illumination of the original image I is regionally adjusted using histogram equalization to generate the re-lit image I_r.

After obtaining the re-lit image I_r, its edge map E_r and its Gabor image G_r of every region are both generated. I_r, E_r, and G_r are used together to localize the eyes. The Gabor image G_r is first used to locate the windows within which the facial features lie. This is because the convolution of Gabor wavelets with an image results in the salient facial features with high magnitudes, such as eyes, nose and mouth. The windows corresponding to eyes are then identified using both the re-lit image I_r and its edge map E_r. After the eyes' windows are detected, the hybrid projection function (HPF) [3] is used to localize the eye positions. This method does not require the knowledge of the orientation of the face nor the illumination of the image. Another advantage of this method is that no initialization is needed. Training data and a training process are not required. The block diagram of the proposed method is shown in Figure 4-2.

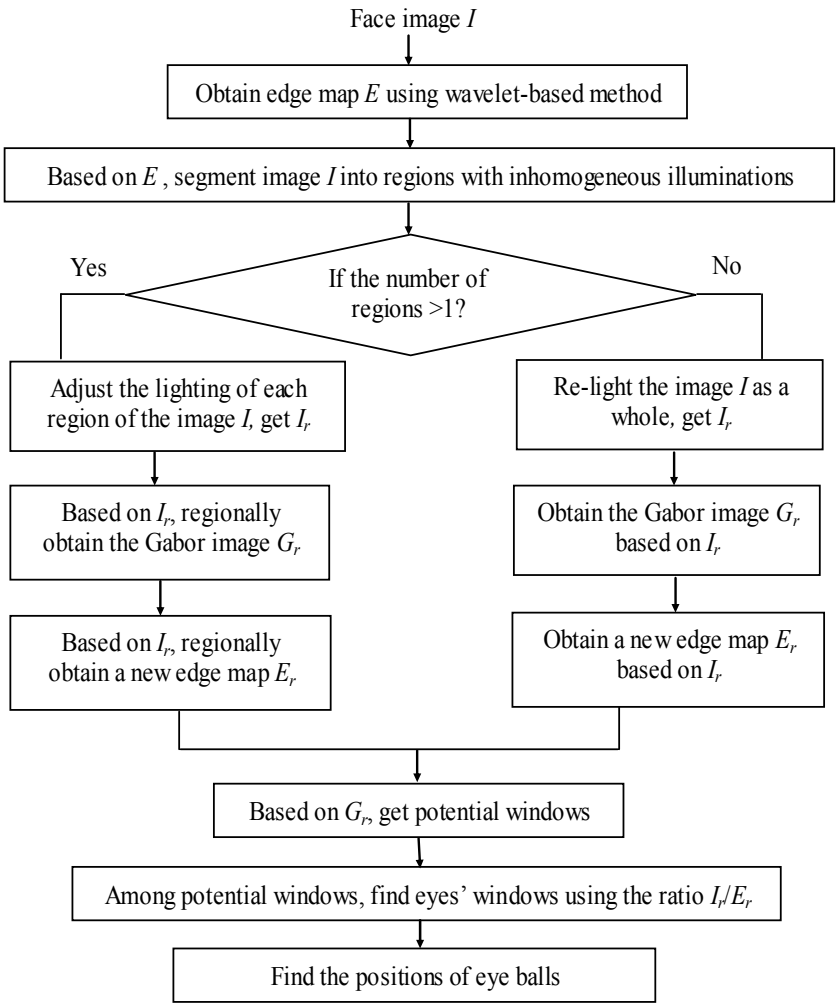


Figure 4-2. Block diagram of the proposed method.

4.2 Related Work

Various techniques have been proposed to detect facial features in face images. In general, two types of information are commonly utilized by these techniques. One is the image appearance of the facial features, which is referred as texture information; and the other is the spatial relationship among the different facial features, which is referred as shape information.

The texture-based methods model the local texture around a given feature point, for example the pixel values in a small region around an eye corner. This type of methods is low-

cost, but not very robust. In [4], the eigenface technique was extended to the facial features, yielding eigeneyes, eigennoses, and eigenmouths. For eye detection, a principal component projective space called "eigeneyes space" is constructed and the query image is compared with the gallery eye images in the eigeneyes space. This method can achieve better eye detection performance than a simple template method since the training samples cover different eye variations in appearance, orientation and illumination conditions. The drawback of it is that the training and test images need to be normalized in size and orientation.

The shape-based methods regard all facial feature points as a shape, which is learned from a set of labeled faces, and try to find the corresponding shape for any unknown face. This type of methods performs well under certain restricted assumptions regarding the head position, the image scale, and the illumination conditions. However, they are computationally demanding. In [5], active contour models (snakes) and an energy minimization approach for edge detection were introduced. These models can be used to detect the face boundary and facial features. This method makes use of the global information to improve the reliability of locating the contour of eyes. In [6], the use of a deformable template for locating human eyes was proposed. This method designs an eye model and the eye position is then obtained through a recursive process. However, this method is only feasible when the initial position of the eye model is placed near the actual eye position. Moreover, the template scheme is associated with problems such as slow convergence and lengthy processing time. To partially solve these problems, the concept of eye corners was introduced to guide the recursive process [7]. In [7][8], the corner detection algorithm was adopted. However, the detection algorithm is based on the edge image, while a good edge image is hard to obtain when the eye image is of relatively low contrast. As a result, the performance of the eye detection

algorithm will be degraded. Active shape model (ASM) [9] is a powerful statistical tool for facial feature detection by shape. However, changes in illumination and facial expressions, and the presence of local minima in optimization form some challenges to this method. In [10], Gabor wavelet features were introduced to ASM for modeling the local image structure.

A number of approaches combining texture- and shape-based methods have also been proposed. In [11], the Gabor wavelets were used to generate a data structure, named the elastic bunch graph, to locate facial features. Active appearance model (AAM) proposed in [12] is a popular shape and appearance model for feature localization. Feature search based on the above models will however become unstable under significant illumination variations. Moreover, the algorithms require a good initialization that is close to the correct solution; otherwise, they are prone to getting stuck in local minima.

Besides the classical methods, many other image-based eye detection techniques have also been reported recently. In [13], Feng *et al.* developed variance projection function (VPF) for locating the landmarks of an eye. It is observed that some eye landmarks are with relatively high contrast, such as the boundary points between the eye white and the eye ball. The located landmarks are then employed to guide the eye detection process. In [14], three cues from the face image were used for detecting eye windows. Each cue indicates the positions of the potential eye windows. The first cue is the face intensity because the intensity of eye regions is relatively low. The second cue is based on the estimated direction of the line joining the centers of the eyes. The third cue is from the response of convolving the processed eye variance filter with the face image. Based on the three cues, a cross-validation process is performed. This process generates a list of possible eye window pairs. For each possible case, variance projection function (VPF) is used for eye detection and

verification. In [3], Zhou *et al.* extended the idea of VPF to the generalized projection function (GPF). Both the integral projection function (IPF) and the variance projection function (VPF) are viewed as special cases of GPF. Another special case of GPF, i.e., the hybrid projection function (HPF), was developed experimentally determining the optimal parameters of GPF. Experiments on three databases showed that IPF, VPF, and HPF are all effective in eye detection. Nevertheless, HPF is better than VPF, while VPF is better than IPF. In [15], a local feature based method was proposed for detecting landmarks from facial images. This method is based on extracting oriented edges and constructing edge maps at two resolution levels. Edge regions with characteristic edge pattern form the landmark candidates. In [16], adaptive boosting (AdaBoost) was used to detect eyes. The boosting algorithm is a method that combines a collection of weak classifiers to form a strong classifier. The AdaBoost algorithm adaptively boosts a sequence of classifiers by dynamically updating the weights according to the errors in the previous learning. In [17], an algorithm for locating the major facial features was developed. This algorithm estimates the parameters of the ellipse which best fits the head view in the image and uses these parameters to calculate the estimated locations of the facial features. It then refines the estimated coordinates of the eyes, mouth, and nose by exploiting the vertical and horizontal projections of the pixels in windows around the estimated locations of the features. In [18], a feature extraction system that relies on the fusion of several facial feature masks derived from multiple feature extractors was proposed. The fusion method is based on the observation that having multiple masks for each feature lowers the probability that all of them are invalid, since each of them produces different error patterns. For each feature, the extracted feature masks are fused together by a dynamic committee machine (DCM) that uses their evaluation to calculate

weights; input image quality in the form of resolution and color quality are used to estimate the gating variables. In [19], an effective facial features detection method for human-robot interaction (HRI) with indoor mobile robot was proposed. The authors suggested a facial feature detection method based on the local image area and the direct pixel-intensity distributions, in which two novel concepts, the directional template for evaluating intensity distributions and the edge-like blob map with multiple strength intensity were proposed. Using the proposed blob map, the locations of major facial features - two eyes and a mouth - can be estimated. Final candidate face region is determined by both the obtained locations of facial features and the weighted correlations with stored facial templates.

Since most of these existing techniques either assume frontal facial views, or without significant facial expressions, or under even illuminations, good performance has been reported. However, in reality, the image appearance of the facial features varies significantly among different individuals. Even for a specific person, the appearance of facial features is easily affected by the lighting conditions, face orientations and facial expressions. Therefore, robust facial feature detection still remains a very challenging task, especially under severely poor illuminations.

In [20], an active infrared (IR)-based approach was proposed to deal with the illumination problem. It is based on the principle of the red-eye effect in flash photographs, utilizing a special IR illuminator and an IR-sensitive CCD for image acquisition. This approach is relatively simple and very effective. However, it requires a special lighting and synchronization scheme. Specialized hardware (an infrared sensitive camera equipped with infrared LEDs) is needed to produce the red eye effect in order to track the pupils.

In this chapter, we propose a novel eye detection method for gray-scale face images that have various illumination conditions. This method consists of four steps: (1) region segmentation based on an edge map obtained via multi-resolution wavelet transform, (2) region-based image re-lighting, (3) extraction of feature regions from the Gabor image constructed from the re-lit image, (4) eye localization based on the edge map of the re-lit image and the re-lit intensity information.

There exist some Gabor-based methods for detecting facial features [10]-[11], [21]-[23]. In [10], Gabor wavelet features were involved in ASM for modeling local image structure. In EBGM [11], facial features were represented with the Gabor *Jets* and the spatial distributions of facial features were captured with a graph structure implicitly. Via the graph structure, only the simple spatial information among the facial features is imposed, whose variation is not modeled directly. In [21], Kalman filtering was first utilized to predict the position of each facial feature in a new image. Then, given the predicted feature positions, the multi-scale and multi-orientation Gabor wavelet matching method was used to detect each facial feature in the vicinity of the predicted locations. [22] suggested an attention-driven approach to feature detection inspired by the human saccadic system. This method computes the Gabor decomposition only on the points of a sparse retinotopic grid and then applies it to eye detection. In [23], an automatic face recognition system based on multiple facial features was described. Each facial feature is represented by a Gabor-based complex vector and is localized by an automatic facial feature detection scheme. Two face recognition approaches, named two-layer nearest neighbor (TLNN) and modular nearest feature line (MNFL) respectively, are then proposed.

All existing Gabor-based methods need to annotate initial feature points in advance and then precise positions of facial features are obtained in the end. The Gabor features are used only for representing the characteristics of the points. Our method is essentially different from the existing methods in that we do not need manual annotation. Gabor features are used for detecting the initial positions of the facial features.

There are some existing facial feature localization methods utilizing an edge map to detect eyes [15][24]. The use of an edge map in our method is significantly different from them in that the edge map is used to segment an image into its differently illuminated regions, rather than to extract the facial feature.

4.3 The Proposed Method

As mentioned earlier, lighting conditions are usually not uniform when imaging a picture. The worst case is the side lighting effect. Normally, an image contrast enhancement method is used to adjust the varying image illumination. However, the conventional histogram equalization method does not work well when extreme conditions are present, especially in the case of side lighting effects. In this chapter, the illumination of the image is first adjusted on a region-by-region basis. The image I is partitioned into n regions with different illuminations based on an edge map E derived from the multi-resolution wavelet transform. The number of regions n can be equal to 1 when the image has even illumination and can be greater than 1 when the image has uneven illumination. Each region is then separately contrast enhanced. The regional re-lighting of the image makes the facial feature details more pronounced.

4.3.1 Regionally Illumination Adjusted Image I_r

As mentioned above, the illumination of an image *I* is first regionally adjusted. The image *I* is partitioned into *n* regions with different illuminations. Region segmentation is based on an edge map of the face image. The edge map is generated using wavelet decomposition.

4.3.1.1 Edge map E obtained using wavelet transform

Image edges involve the high frequency components of an image. And often, edges occur at different resolutions; both strong edges and weak edges exist in the same image. It is appropriate to extract edges at different scales or resolutions.

Wavelet decomposition of an image provides a good solution to obtaining the edge map. Here we use the redundant wavelet transform. The decomposition procedure for a redundant wavelet transform is different from the traditional one in that the scaling of the wavelet is not achieved by sub-sampling the image at each step, but rather by an up-sampling of the filters. The four wavelet sub-bands at scale *j* are of the same size as the original image, and all filters used at scale *j* are up-sampled by a factor of 2^j (padding $2^j - 1$ zeros) compared with those at scale zero.

The edges of an image are full of high frequency information that scatters into several scales or resolutions. In order to take advantage of the multi-resolution property of wavelet transforms, two corresponding sub-bands at adjacent resolutions are multiplied so as to enhance image edges and suppress noise. This is based on the fact that edge structures are present at each scale while noise decreases rapidly along the scales. For more details, please refer to [25]. Figure 4-3a, 4-3b and 4-3c show the original images with different illuminations and Figure 4-3d, 4-3e and 4-3f show their edge maps respectively.

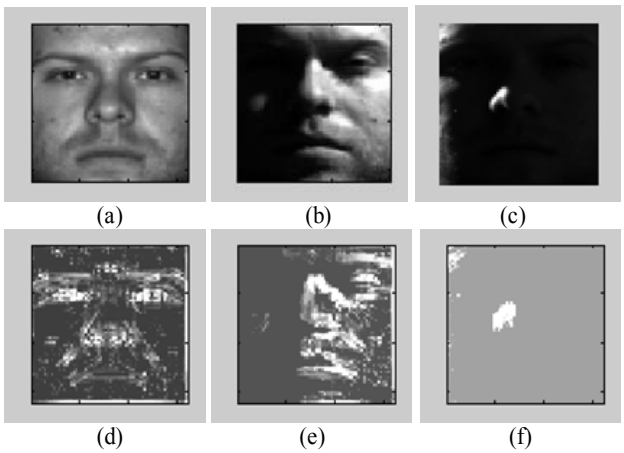


Figure 4-3. Edge maps.

(a), (b) and (c): original images; (d), (e) and (f): edge maps

4.3.1.2 Region segmentation

Now we use our observation that for poor illumination, the edges have extremely weak intensity strengths. Actually the lack of light or over lighting both weakens the real edges. Therefore, we examine the edge map. If some edges in the edge map are weak, it means the illumination in this region may be either over-lit or under-lit. We differentiate these regions from the well-lit regions. In Figure 4-4, the segmented regions of the three differently illuminated images are shown using blue lines. Here, we should note that the number of regions may be greater than or equal to 1. For an evenly illuminated image, the number of regions is only 1. For an unevenly illuminated image, the number of regions is greater than 1. Histogram equalization is then applied to each region separately. Figure 4-4a, 4-4b and 4-4c show the region segmentation based on the edge maps. Figure 4-4d, 4-4e and 4-4f show the regions on the original intensity images.

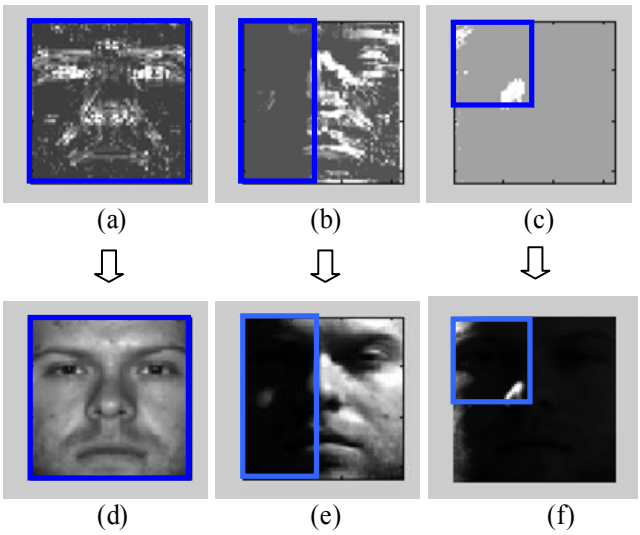


Figure 4-4. Segmented regions.

(a), (b) and (c): edge maps; (d), (e) and (f): regions of the original images

4.3.1.3 Regionally re-lit intensity image $\mathbf{I_r}$

The re-lit (illumination adjusted) image I_r is obtained by applying histogram equalization to each region separately. Figure 4-5a, 4-5c and 4-5e show the original images. Figure 4-5b, 4-5d and 4-5f show the resulting regionally re-lit images, respectively.

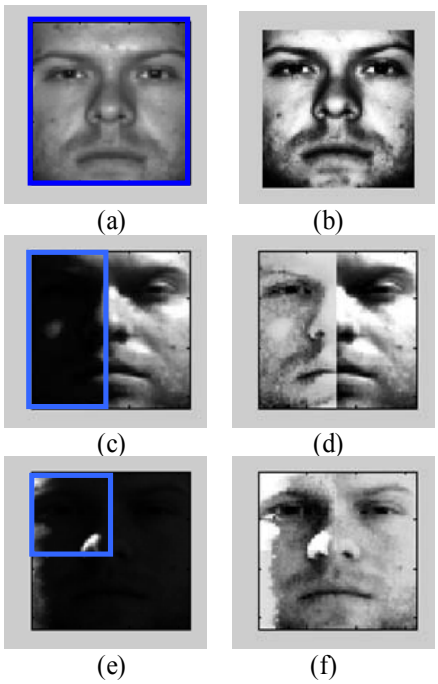


Figure 4-5. Regionally re-lit intensity image I_r (b), (d), and (f).

4.3.2 Localizing the Eyes' Features

4.3.2.1 Gabor wavelet convolution and regionally obtained Gabor image G_r

Gabor images are the results of convolving the face image with a set of Gabor wavelets. A Gabor image exhibits high magnitudes on the salient facial features, such as eyes, nose and mouth. We thus use Gabor images to locate windows within which the facial features lie.

The 2-D Gabor wavelets are defined as follows:

$$\psi_j(\vec{x}) = \frac{\|\vec{k}_j\|^2}{\sigma^2} \exp\left(-\frac{\|\vec{k}_j\|^2 \|\vec{x}\|^2}{2\sigma^2}\right)\left[\exp(i\vec{k}_j\vec{x}) - \exp\left(-\frac{\sigma^2}{2}\right)\right] \tag{4-1}$$

where

$$\vec{k}_j = k_m e^{i\varphi_n} \tag{4-2}$$

$$k_m = \frac{0.5\pi}{(\sqrt{2})^m} \quad \varphi_n = n\frac{\pi}{8} \tag{4-3}$$

In this chapter, three different scales' Gabor wavelets (Figure 4-6) of horizontal orientation with $m \in \{0,1,2\}$ and $n = 0$ are used to derive the Gabor wavelet convolved images.

Figure 4-6. 3 horizontal Gabor wavelets.

The Gabor wavelet representation $O_j(\vec{x})$ of an image is the convolution of the image $I(x, y)$ with a family of Gabor wavelets $\psi_j(\vec{x})$,

$$O_j(\vec{x}) = I(\vec{x}) * \psi_j(\vec{x}) \qquad j = 1 \to 3 \tag{4-4}$$

where $(\vec{x}) = (x, y)$, and * denotes the convolution operator.

Figure 4-7a shows the original image and Figure 4-7b shows the three-scale Gabor convolved images. The three images are then multiplied and thresholded to highlight the high responses and suppress noise. Thus,

$$G_r(\vec{x}) = O_1(\vec{x}) \times O_2(\vec{x}) \times O_3(\vec{x}) \quad (4\text{-}5)$$

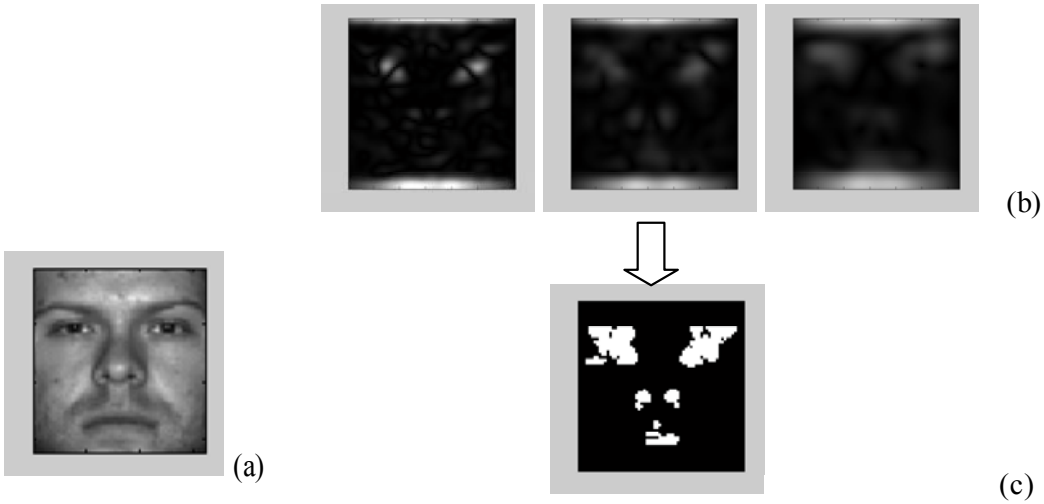


Figure 4-7. Gabor convolved image.

Figure 4-7c is the Gabor image G_r obtained by multiplying the three-scale convolved images and then thresholding the product. Normally, we get high magnitudes at the images' upper and bottom outer edges. As they are not associated with any facial feature, they are removed.

For illustration purpose, in Figure 4-7, a well-lit image which has one illumination region (i.e., n=1) is used. G_r is obtained using the whole image. However, if I has more than one region (n>1), then Gabor image G_r is obtained on a region by region basis, i.e., by applying the above Gabor convolution to each region separately. Figure 4-8a and 4-8d show the original images. Figure 4-8b and 4-8e show the Gabor images obtained using the whole image. Figure 4-8c and 4-8f show the regionally obtained Gabor images G_r.

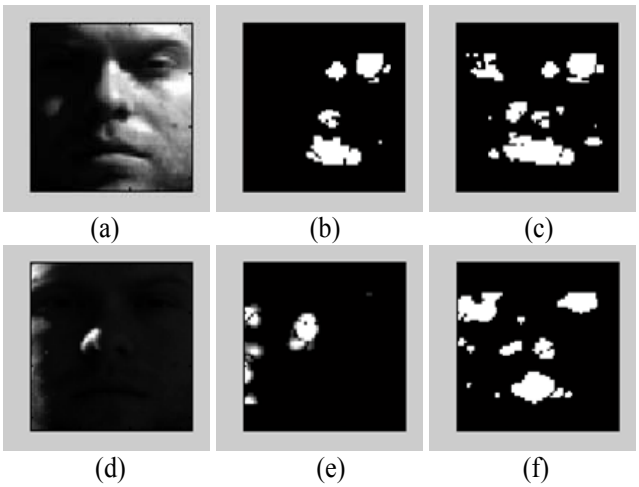


Figure 4-8. Regionally obtained Gabor image G_r (c), (f).

4.3.2.2 Localization of facial features' windows

The Gabor image G_r is used to locate the windows within which different facial features lie. The Gabor image consists of the high response concentrations presumed to contain facial landmarks. The regions of G_r with high magnitudes are clustered, based on their positions to form isolated windows. These windows are the potential windows presumed to contain the facial features. In order to localize the separate windows containing the facial features, the Gabor image G_r is first partitioned into small blocks so as to filter out the scattered isolated high magnitude points and to speed up the following clustering process. The blocks with magnitudes higher than a threshold are marked as candidate blocks. The candidate blocks are clustered based on their coordinates. The eight nearest neighbours method is used to connect the high magnitudes' blocks. The connected blocks form the potential features' windows. We label these windows as $w_1, w_2, ..., w_n$. The clustering method is described in Figure 4-9. The potential windows are shown in Figure 4-10.

1. Partition the Gabor image G_r into blocks, e.g., 4×4.
2. For each block, compute the average magnitude:
   If the average is larger than a pre-defined threshold, merge it with the adjacent blocks with higher magnitudes using the eight nearest neighbours method.
3. The connected blocks form isolated regions. The isolated regions form the potential features' windows.

Figure 4-9. Clustering algorithm.

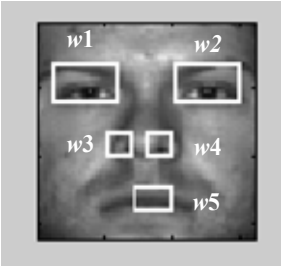


Figure 4-10. Potential features' window.

4.3.2.3 Regionally obtained new edge map E_r

Similarly to G_r, the regionally obtained new edge map E_r is obtained by applying the above wavelet transform method to different regions of I_r separately. Figure 4-11a and 4-11d show the original images. Figure 4-11b and 4-11e show the edge maps obtained using the whole image. Figure 4-11c and 4-11f show the regionally obtained edge maps I_r.

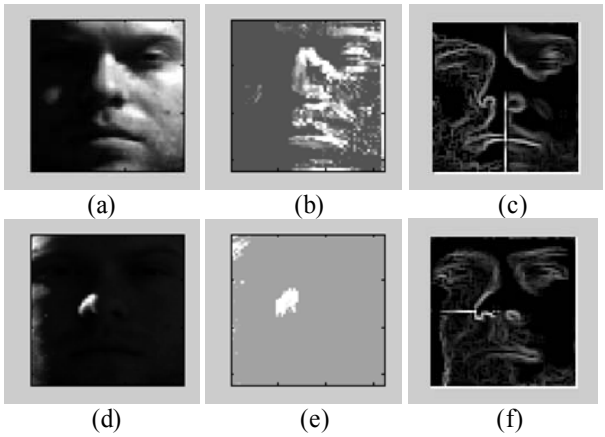


Figure 4-11. Regionally obtained new edge map E_r (c), (f).

4.3.2.4 Determining the "eyes" windows

The pixel and edge intensity information are used to verify the existence of eyes in the image. To identify the windows in G_r that contain the eyes, we rely on I_r and E_r. Compared with the other features, the eyes have the lowest average gray-scale intensity since eyes normally have low intensity, but the highest average edge intensity since there are a lot of edge points near the eyes' positions. Thus, we compute the ratio of the average gray-scale intensity to the average edge intensity for each window i to obtain

$$Eye\_window = \arg\min_i \frac{\bar{I}_r^{\,i}}{\bar{E}_r^{\,i}} \tag{4-6}$$

where $\bar{I}_r^{\,i}$ is the average intensity value of window i, and $\bar{E}_r^{\,i}$ is the average edge intensity value of window i. The eyes' windows are chosen as the two windows with the two lowest ratios. Figure 4-12 shows the eyes' windows.

Figure 4-12. Eyes' windows.

4.3.2.5 Locating eye positions

After locating the eyes' windows, the eye position is detected using the hybrid projection function (HPF) [3]. The hybrid projection function is a combination of the integral projection function (IPF) and the variance projection function (VPF).

Suppose $I(x,y)$ is the intensity of an image, the vertical integral projection $IPF_v(x)$ and the horizontal integral projection $IPF_h(y)$ of $I(x,y)$ in the intervals $[y_1,y_2]$ and $[x_1,x_2]$ can be defined respectively as

$$IPF_v(x)=\int_{y_1}^{y_2} I(x,y)dy \tag{4-7}$$

$$IPF_h(y)=\int_{x_1}^{x_2} I(x,y)dx \tag{4-8}$$

Usually the mean vertical and horizontal projections are used, which can be defined respectively as

$$IPF_v'(x)=\frac{1}{y_2-y_1}\int_{y_1}^{y_2} I(x,y)dy \tag{4-9}$$

$$IPF_h'(y)=\frac{1}{x_2-x_1}\int_{x_1}^{x_2} I(x,y)dx \tag{4-10}$$

Although IPF is the most commonly used projection function, it cannot well reflect the variation in the image. Thus, VFP was proposed by Feng *et al.* [13]. Suppose $I(x,y)$ is the intensity of an image, the vertical variance projection $VPF_v(x)$ and the horizontal variance projection $VPF_h(y)$ of $I(x,y)$ in the intervals $[y_1,y_2]$ and $[x_1,x_2]$ can be defined respectively as

$$VPF_v(x)=\frac{1}{y_2-y_1}\int_{y_1}^{y_2}[I(x,y)-IPF_v'(x)]dy \tag{4-11}$$

$$VPF_h(y) = \frac{1}{x_2 - x_1}\int_{x_1}^{x_2}[I(x,y) - IPF_h'(y)]dx \quad (4\text{-}12)$$

Because IPF considers the mean of intensity while VPF considers the variance of intensity, Zhou *et al.* [3] proposed a new projection function, the generalized projection function (GPF) to combine them together. Suppose $I(x,y)$ is the intensity of an image, the vertical generalized projection $GPF_v(x)$ and the horizontal generalized projection $GPF_h(y)$ of $I(x,y)$ in the intervals $[y_1, y_2]$ and $[x_1, x_2]$ can be defined respectively as

$$GPF_v(x) = (1-\alpha)IPF_v'(x) + \alpha VPF_v(x) \quad (4\text{-}13)$$

$$GPF_h(y) = (1-\alpha)IPF_h'(y) + \alpha VPF_h(y) \quad (4\text{-}14)$$

where $0 \le \alpha \le 1$ is used to control the relative contributions of IPF and VPF. It is obvious that both IPF and VPF are special cases of GPF where $\alpha = 0$ or 1, respectively. Another special case of GPF, where $\alpha = 0.6$ is proposed by Zhou *et al.* [3] to detect eyes. In Figure 4-13, we show the eyes' positions obtained using the hybrid projection function.

Figure 4-13. Eyes' positions.

4.4 Experimental Results

To evaluate the performance of the proposed eye detection method, we test it on the Yale Face Database B [26]. Yale B is a representative face image database used widely to evaluate face recognition techniques. The database consists of 10 subjects. Each subject has 9 poses and 64 different illumination conditions. In order to examine the effectiveness of the

proposed method, we apply it to the images of different persons with different illuminations and different poses.

The illumination conditions in the Yale face database B images consists of five subsets according to the light-source directions (azimuth and elevation): Subset 1 (angle < 12 degrees from optical axis), Subset 2 (20 < angle < 25 degrees), Subset 3 (35 < angle < 50 degrees), Subset 4 (60 < angle < 77 degrees), and Subset 5 (others). The range of poses is from the half left side profile to the half right side profile. Totally, 730 images are used to test the method.

Figure 4-14 shows the detection results on an image with the side lighting effect: (a) is the original image I; (b) is the original Gabor image G; (c) is the original edge map E; (d) is the regionally re-lit image I_r; (e) is the regionally obtained Gabor image G_r; (f) is the regionally obtained new edge map E_r; (g) shows the potential features' windows; (h) shows the eyes' positions. Figure 4-15 shows the detection results on an image that is almost all dark: (a) is the original image I; (b) is the original Gabor image G; (c) is the original edge map E; (d) is the regionally re-lit image I_r; (e) is the regionally obtained Gabor image G_r; (f) is the regionally obtained new edge map E_r; (g) shows the potential features' windows; (h) shows the eyes' positions.

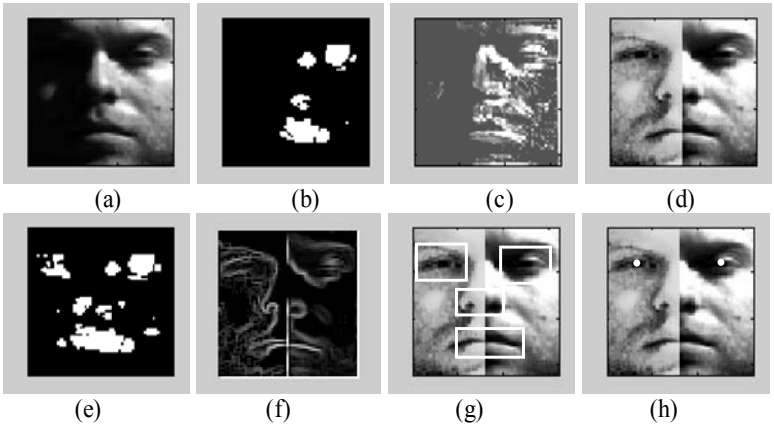


Figure 4-14. Detection on the image with side lighting effect.

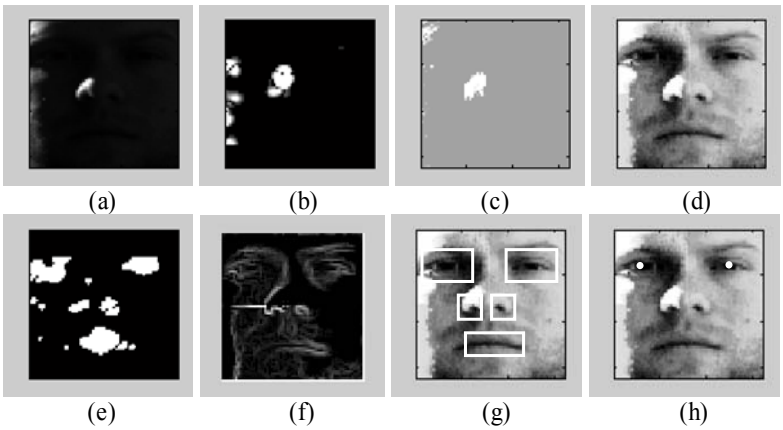


Figure 4-15. Detection on the image that is almost dark.

More experimental results are shown in Figure 4-16 - Figure 4-20. Figure 4-16 shows the detection results on images with different poses. Figure 4-17 shows the detection results on images of different persons. Figure 4-18 shows the eye windows on images of different persons with different poses. Figure 4-19 shows the corresponding eye positions.

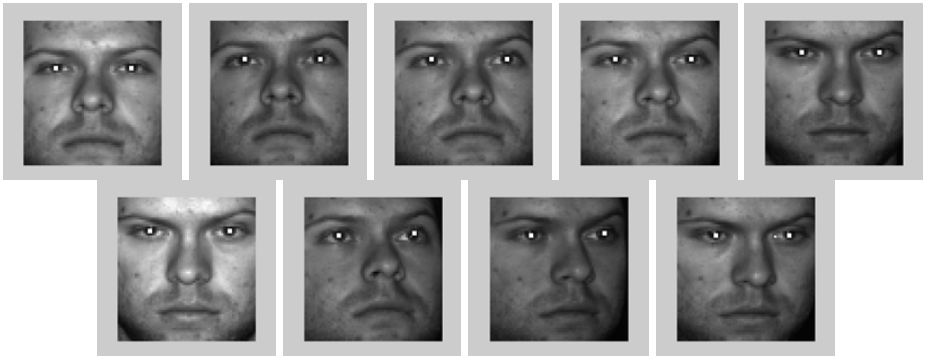


Figure 4-16. Detection on images in different poses.

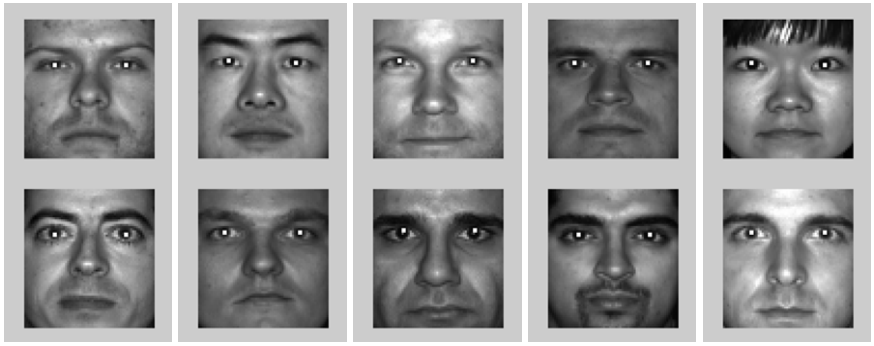


Figure 4-17. Detection on images of different persons.

Figure 4-18. Eye windows.

Figure 4-19. Eye positions.

More detection results on differently illuminated images are shown in Figure 4-20.

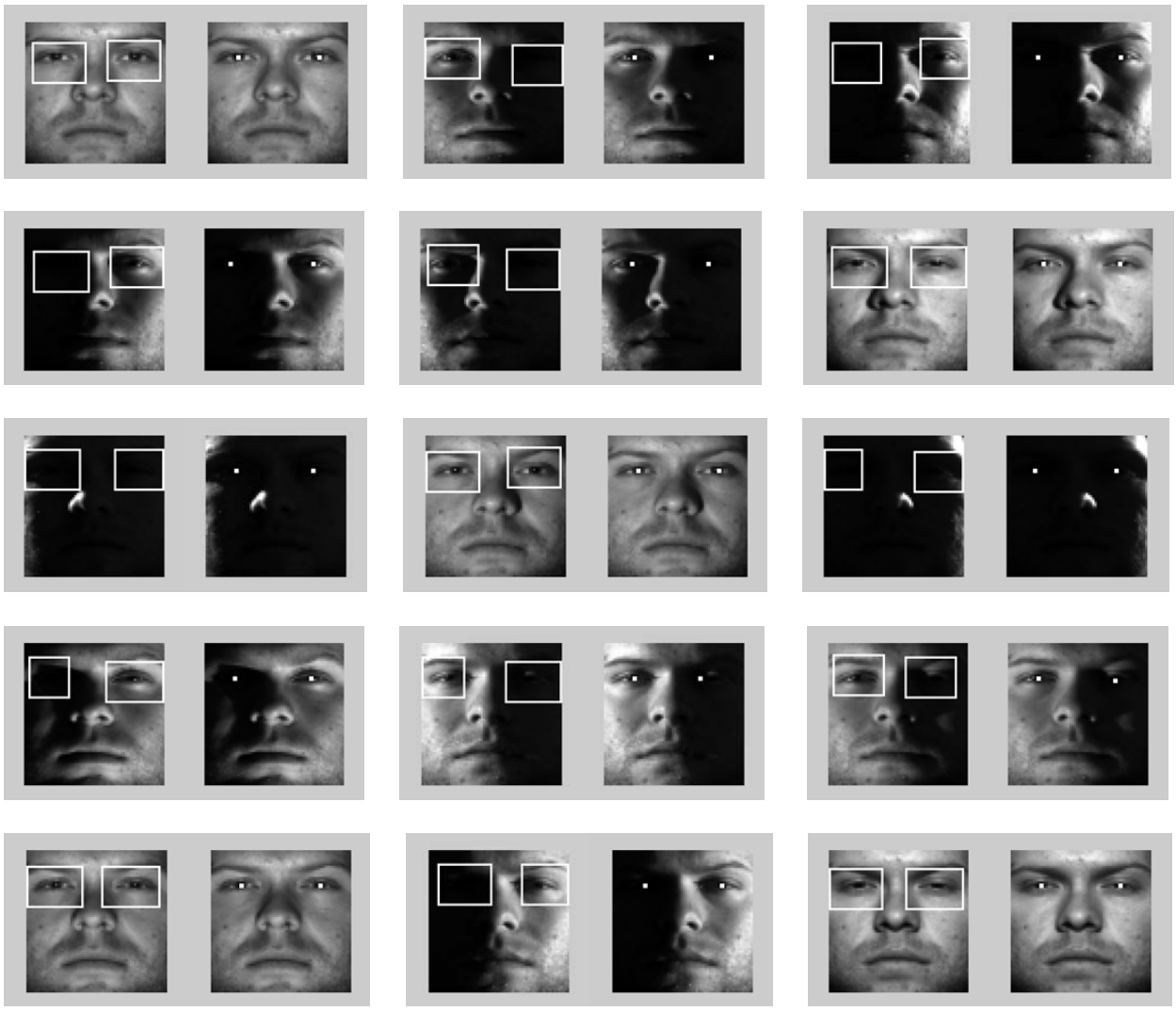


Figure 4-20. Detection results on differently illuminated images.

We regard the detection scheme as successful if the eye position is detected within the iris. The correct detection rate on the pose subset is 96.6% (90 images in total). The correct detection rate on the illumination subset is 90.6% (640 images in total).

4.5 Conclusions

In this chapter, we propose an efficient method to find the windows containing facial features and detect the approximate positions of the eyes. The regionally re-lit gray-scale image I_r, the regionally obtained Gabor image G_r, and the regionally obtained edge map E_r are used together to facilitate this task. This method does not require prior knowledge about face orientation and illumination strength. Other advantages are that no initialization and training process are needed. The experimental results show that this method works well for face images of different people and under various conditions such as changes in lighting and pose.

4.6 References

[1] M. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34-58, 2002.

[2] P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.

[3] Z. Zhou and X. Geng, "Projection functions for eye detection," *Pattern Recognition*, vol. 37, pp. 1049-1056, 2004.

[4] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 84-91, 1994.

[5] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *Proc. of International Conference on Computer Vision*, pp. 259-268, 1987.

[6] A. Yuille, D. Cohen, and P. Hallinan, "Feature extraction from faces using deformable templates," *Proc. of International Conference on Computer Vision and Pattern Recognition*, pp. 104-109, 1989.

[7] K. Lam and H. Yan, "Locating and extracting the eye in human face images," *Pattern Recognition*, vol. 29, no. 5, pp. 771-779, 1996.

[8] X. Xie, R. Sudhakar, and H. Zhang, "On improving eye feature extraction using deformable templates," *Pattern Recognition*, vol. 27, no. 6, pp. 791-799, 1994.

[9] A. Lanitis, C. Taylor, and T. Cootes, "Automatic face identification system using flexible appearance models," *Image and Vision Computing*, vol. 13, pp. 393-401, 1995.

[10] F. Jiao, S. Li, H. Shum, and D. Schuurmans, "Face alignment using statistical models and wavelet features," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 321-327, 2003.

[11] L. Wiskott, J. Fellous, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 775-779, 1997.

[12] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681-685, 2001.

[13] G. Feng and P. Yuen, "Variance projection function and its application to eye detection for human face recognition," *Pattern Recognition Letter*, vo. 19, pp. 899-906, 1998.

[14] G. Feng and P. Yuen, "Multi-cues eye detection on gray intensity image," *Pattern Recognition*, vol. 34, pp. 1033-1046, 2001.

[15] Y. Gizatdinova and V. Surakka, "Feature-based detection of facial landmarks from neutral and expressive facial images," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, vol. 28, no. 1, pp. 135-139, 2006.

[16] C. Park, J. Kwak, H. Park, and Y. Moon, "An effective method for eye detection based on texture information," *Proc. of International Conference on Convergence Information Technology*, pp. 586-589, 2007.

[17] A. Alattar and S. Rajala, "Facial features localization in front view head and shoulders images," *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, pp. 3557-3560, 1999.

[18] S. Ioannou, M. Wallace, K. Karpouzis, A. Raouzaiou, and S. Kollias, "Combination of multiple extraction algorithms in the detection of facial features," *Proc. of IEEE International Conference on Image Processing*, vol. 2, pp. 378-381, 2005.

[19] T. Lee and S. Park, and M. Park, "A new facial features and face detection method for human-robot interaction," *Proc. of IEEE International Conference on Robotics and Automation*, pp. 2063- 2068, 2005.

[20] Z. Zhu and Q. Ji, "Robust real-time eye detection and tracking under variable lighting conditions and various face orientations," *Computer Vision and Image Understanding*, vol. 98, pp. 124-154, 2005.

[21] Z. Zhu and Q. Ji, "Robust pose invariant facial feature detection and tracking in real-time," *Proc. of International Conference on Pattern Recognition*, vol. 1, pp. 1092-1095, 2006.

[22] F. Smeraldi and J. Bigun, "Facial feature detection by saccadic exploration of the Gabor decomposition," *Proc. of the International Conference on Image Processing*, vol. 3, pp. 163-167, 1998.

[23] R. Liao and S. Li, "Face recognition based on multiple facial features," *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 239-244, 2000.

[24] J. Song, Z. Chi, and J. Liu, "A robust eye detection method using combined binary edge and intensity information," *Pattern Recognition*, vol. 39, pp. 1110-1125, 2006.

[25] S. Du and R. Ward, "Adaptive region-based image enhancement method for face recognition under varying illumination conditions," *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 353-356, 2006.

[26] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, 2001.

CHAPTER 5 FACE RECOGNITION UNDER POSE VARIATIONS: A SURVEY[4]

5.1 Introduction

Existing work in face recognition has demonstrated good recognition performance on frontal, expressionless views of faces taken under controlled lighting conditions. However, a practical face recognition system must also work under other imaging conditions, such as different face poses.

In this chapter, we present a review of face recognition techniques under different pose conditions. In Section 5.2, we classify the face recognition algorithms that address pose variations into three categories: (1) the invariant features extraction-based approach, (2) the multiview-based approach, and (3) the 3D range image-based approach. We describe the major features of each category and summarize the methods that belong to each approach. In Section 5.3, future research challenges are identified. Section 5.4 concludes the chapter.

5.2 Overview of Pose-invariant Face Recognition Algorithms

Past studies in the field of automatic face recognition have revealed that the biggest challenge is to reliably recognize people in the presence of image/object variations that occur naturally in our daily life. One of the most common variations is in head pose. Handling head pose variations is extremely important in many practical applications. When the face is rotated in the image plane, it can be normalized by detecting at least two facial features. However, when the face is subjected to in-depth rotation, geometrical normalization is not

[4] A version of this chapter has been published in Journal of the Franklin Institute, Shan Du and Rabab Ward, "Face Recognition under Pose Variations," vol. 343, no. 6, pp. 596-613, 2006.

possible. The situation becomes even worse when lighting differences, occlusions and self-shadowing of facial features are present.

There are different algorithms for tackling the pose variation problem. In this chapter, we categorize them as follows: (1) the invariant features extraction-based approach, (2) the multiview- based approach, and (3) the 3D range image- based approach.

The invariant features extraction-based approach records some features in a face image that do not vary under pose changes, such as color or geometric invariants [1]-[4]. In this approach, the appearance-based method is to learn some suitable subspace/manifold representing the extent of the variations. The model-based method uses 2D deformable models to capture variations in pose.

The multiview-based approach stores multiview images in the gallery to deal with the pose variation problem (real multiview images) or to synthesize new view images from a given image (synthesized multiview images). Recognition is then performed using both the given image and the synthesized images. The synthesis algorithms normally use prior knowledge of faces to synthesize new views from one view. A generic 3D model of the human face can be used to predict the appearance of a face under different pose parameters [5]-[7]. Once a 2D face image is texture mapped onto a 3D model, the face can then be treated as a traditional 3D object in computer graphics and undergo 3D rotation. Another method that uses prior knowledge of faces to synthesize new views from one view is the linear object classes method [8][9]. This method represents prior face knowledge using 2D views of prototype faces. These miltiview-based methods, which make use of prior class information, have so far been the most successful and practical techniques for solving the pose variation problem.

Face recognition from 3D range images is the third approach being actively studied by researchers. Since a face is inherently a 3D object, a good solution would be to use information about the 3D structure of a face. A 3D range image contains the depth structure of the object, and can represent a 3D shape explicitly and compensate for the lack of depth information in a 2D image. The 3D shape is invariant to changes in color or to changes in reflectance properties resulting from variations in ambient lighting. Because the shape of a face is not affected by changes in lighting or pose, the 3D face recognition approach has the potential to improve performance when such changes occur.

The methods within each of the above three approaches can be further classified into subclasses according to their respective characteristics. We describe and summarize each of these approaches below.

5.2.1 Invariant Features Extraction-based Approach

This approach extracts features that are invariant to pose changes. We categorize this approach into appearance-based algorithms and geometric model-based algorithms. In the appearance-based algorithms, an image is considered as a high-dimensional vector, i.e., a point in a high-dimensional vector space. Statistical techniques are used to analyze the distribution of the object image vectors in the vector space, and derive an efficient and effective representation (feature space). Given a probe image, the similarity between the gallery images and the probe image is then carried out in the feature space. The geometric model-based algorithms use 2D deformable models to capture the pose variations. Recognition is carried out by comparing the specified model of the probe image with those of the gallery images. The methods that are representative of this approach are summarized in Table 5-1.

5.2.1.1 Appearance-based algorithms

In the appearance-based algorithms, images are represented as vectors, i.e., as points in a high-dimensional vector space. For example, an $m \times n$ 2D image can be mapped to a vector $x \in R^{mn}$ by concatenating each row or column of the image. Based on the different statistical methods, a set of basis vectors are obtained for the high-dimensional face vector space. By projecting the face vector to the basis vectors, the projection coefficients are used as the feature representation of each face image. The matching score between the probe face image and a gallery image is calculated using their coefficients' vectors. The larger the matching score, the better the match.

Eigenface [10] is a typical method belonging to this class. The Eigenface method uses principal component analysis (PCA) for dimensionality reduction to find the vectors that best account for the distribution of face images within the entire image space. The key procedure in PCA is based on the Karhumen-Loeve transformation [11]. The basis vectors are defined as the eigenvectors of the scatter matrix S_T,

$$S_T = \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T \qquad (5\text{-}1)$$

where $x_1, x_2, ..., x_N$ are the images in the training set, and μ is the average image of this training set. The eigenvectors corresponding to the d largest eigenvalues of S_T compose the transformation matrix W_{PCA}.

Since PCA constructs the face space without using face class information, it does not perform well on the pose variation problem. To improve it, the Fisherface method [3] has been proposed to find an efficient way to represent the face vector space. Exploiting face class information can be helpful in performing pose-invariant tasks.

The Fisherface algorithm is derived from the Fisher linear discriminant/linear discriminant analysis (FLD/LDA) [3], which uses class-specific information. By defining different classes with different statistics, the images in the training set are divided into the corresponding classes. Then, techniques similar to those used in the Eigenface algorithm are applied. The Fisherface algorithm results in a higher accuracy rate in recognizing faces, compared with the Eigenface algorithm.

The linear discriminant analysis (LDA) finds a transform, W_{LDA}, such that

$$W_{LDA} = \arg\max_{w} \frac{W^T S_B W}{W^T S_w W} \tag{5-2}$$

where S_B is the between-class scatter matrix and S_w is the within-class matrix, defined as

$$S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T \tag{5-3}$$

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \tag{5-4}$$

where N_i is the number of training samples in class *i*, *c* is the number of distinct classes, μ_i is the mean vector of samples belonging to class *i* and X_i represents the set of samples belonging to class *i*. Though the original Fisherface paper [3] did not consider the pose problem, the authors pointed out that this method can be extended to handle pose variations.

In [13], a feature-based pose estimation and face recognition system that uses 2D Gabor features as local feature information was proposed. The difference between this system and existing ones lies in its simplicity and intelligent sampling of local features. A learning algorithm that tries to learn the importance of each uniformly sampled grid point is employed. To select the most informative grid points for the task, the jet response of each grid point is used as an isolated feature vector for a given test image. Using this feature vector only, the

performance of the system is calculated. Higher performance means that the selected point contains useful information, and thus deserves a higher weight, proportional to its performance. Applying this scheme to each grid point on the sampling lattice, weights for each point are obtained. In [13], the authors compared the performance of the system with the standard modular Eigenfaces approach, and showed that local feature-based approach improves the performance of both pose estimation and face recognition.

5.2.1.2 Geometric model-based algorithms

The geometric model-based face recognition scheme aims at constructing a model of the human face that is capable of capturing facial variations. Prior knowledge of the human face is used in the design of this model.

Model-based schemes usually involve three steps: (1) constructing the model, (2) fitting the model to the given face image, and (3) using the parameters of the fitted model as the feature vector to calculate the similarity between the probe face and the gallery faces in the database to perform the recognition.

In the elastic bunch graph matching method (EBGM) [1], faces are represented as graphs, with nodes positioned at fiducial points (such as the eyes, the tip of the nose, some contour points, etc.), and edges labeled with 2D distance vectors. Each node contains a set of 40 complex Gabor wavelet coefficients, including both phase and magnitude, known as a jet. Thus, the geometry of an object is encoded by the edges, while the gray-values distribution is patch-wise encoded by the nodes (jets). Face recognition is based on labeled graphs matching. Classification experiments have shown that this approach can handle changes in expression and pose, as well as small changes in lighting conditions.

The active shape model (ASM) [12] is a statistical flexible model built by learning patterns of variability from a training set of correctly annotated images. This model takes both class specificity and variability into account. Each object is represented by a set of points. The points can represent the boundary, internal features, or even external features. Points are placed in the same way on each object of a training set of examples. By examining the statistics of the positions of the labeled points, a point distribution model (PDM) is derived. The model gives the average positions of the points, and has a number of parameters that control the main modes of variation found in the training set. A compact parameterized description of shape for any instance of a face is thus provided, and can be used in a multi-resolution search to locate the features in new images. Gray-level appearance is modeled using flexible gray-level models analogous to the shape model. The primary description is provided by a shape-free gray-level model of the whole face. Local gray-level models, attached to points on the shape model, are also used to make the ASM search more robust and improve person identification in the presence of partial occlusion. Given such a model and an image containing an example of the object modeled, image interpretation involves choosing values for each of the parameters to find the best fit of the model to the image. For an input probe image, all three types of information, including extracted shape parameters, shape-free image parameters, and local profiles, are used for classification.

Active appearance model (AAM) [2] is an integrated statistical model that models and uses both shape and gray-level information. Matching the model to an image involves finding the model parameters that minimize the difference between the image and the synthesized model example, projected onto the image. The potentially high number of parameters makes this a difficult problem. In [2], an efficient direct optimization approach that matches shape

and texture simultaneously was proposed, resulting in a rapid, accurate, and robust optimization. For rapid matching, the authors pre-computed the derivatives of the residual of the match between the model and the target image and used them to compute the update steps in an iterative matching fashion.

Table 5-1. Face recognition methods using invariant features.

Method	Database	Image Size	Subject No.	Image No. per Subject	Image Type	Training Image No. per Subject	Recognition Rate (%)
Fisherface [3]	Harvard	n/a	5	66	Gray-level 2D	30	n/a
Feature selection [13]	ESRA, CVL	32×32 32×32	20 113	54 6	Gray-level 2D	18 3	97.3% 54.1%
Elastic Bunch Graph Matching (EBGM) [1]	FERET, Bochum	256×384 128×128	250 108	~6 ~4	Gray-level 2D	1 1	varying 9-98% varying 88-94%
Active Shape Model (ASM) [12]	Self-defined	n/a	30	23	Gray-level 2D	10	92%
Active Appearance Model (AAM) [2]	Self-defined	n/a	n/a	200 in total	2D	100 in total	n/a

5.2.2 Multiview-based Approach

To deal with the pose variation problem, the multiview-based approach stores multiview images in the gallery (real multiview images) or synthesizes new view images from a given image (synthesized multiview images). Recognition is then performed using both the given image and the synthesized images. The representative methods of this approach are summarized in Table 5-2.

5.2.2.1 Multiple real image-based algorithms

Having multiview images stored in the gallery is one strategy for dealing with the pose variation problem. It is a direct extension of the frontal face recognition technique. Compared

with the invariant features extraction-based approach, the multiview-based approach should be able to achieve better results when the out-of-plane rotation angle is large. One such algorithm is in [14], which uses a template-based correlation matching scheme. In this work, after a pose estimation step, the algorithm geometrically aligns the probe image to the candidate pose in the gallery images. The alignment is first carried out via a 2D affine transformation based on the automatically determined locations of three facial feature points (eyes and nose), then the optical flow method is used to refine the alignment. Recognition is performed by computing normalized correlation scores. A good recognition rate of 98% is reported on a database of 62 people containing 10 testing and 15 modeling views per person imaged in a number of poses, ranging from -30° to 30° (yaw) and from -20° to 20° (pitch). The main limitations of this method are that (1) many different views per person are needed in the database, and (2) no lighting variations or facial expressions are allowed.

In [15], the popular Eigenface approach was extended to handle multiple views. The performance of a parametric eigenspace (computed using all views from all subjects) is experimentally compared with the performance of view-based eigenspaces (separate eigenspaces for each view). In the experiments, the view-based eigenspaces outperform the parametric eigenspace. Along with a modular eigenspace technique and an automatic feature extraction technique, the system achieves a recognition rate of 95% on a database of 7562 images of approximately 3000 individuals.

A view-based statistical method based on a small number of 2D statistical models (AAM) was proposed in [16]. Unlike most existing methods that can only handle images with rotation angles up to certain degrees, the authors stated that their method can even handle profile views in which many features are invisible. To deal with such large pose variations,

sample views at 90^o (full profile), 45^o (half profile), and 0^o (frontal view) are required. In this method, separate active appearance models are trained for profile, half profile and frontal views. Given a single image of a new person, all the models are used to match the image, and estimation of the pose is achieved by choosing the best fit.

The illumination cone-based method was proposed in [18][19] to handle both pose and illumination problems in face recognition. This method exploits the fact that the set of images of an object in a fixed pose, but under all possible illumination conditions, forms a convex cone in the space of images. Using a small number of training images of each face taken with different lighting directions, the shape and albedo of the face can be reconstructed. In turn, this reconstruction serves as a generative model that can be used to synthesize images of the face under novel poses and illumination conditions. The pose space is then sampled and, for each pose, the corresponding illumination cone is approximated by a low-dimensional linear subspace whose basis vectors are estimated using the generative model. To handle variations due to rotation, the generalized-bas-relief (GBR) ambiguity needs to be completely resolved and then the Euclidean 3D shape reconstructed. The authors proposed a pose- and illumination-invariant face recognition method based on building illumination cones at each pose for each person. Though this is a good idea conceptually, in practice it is too expensive to implement. The authors suggested many ways of speeding up the process, including first sub-sampling the illumination cone and then approximating the sub-sampled cone with an 11D linear subspace. Experiments in building an illumination cone and in 3D shape reconstruction based on seven training images per class are reported. Test results show that the method performs almost without error, except under the most extreme lighting directions.

In [20], a framework for pose-invariant face recognition was proposed. It uses parametric linear subspace models as stored representations of known individuals. Each model can be fitted to an input, resulting in faces of known people whose head pose is aligned to the input face. The model's continuous nature enables the pose alignment to be very accurate, improving the recognition performance, while its generalization to unknown poses enables the models to be compact. Recognition systems with two types of parametric linear model are compared using a database of 20 persons. The experimental results demonstrate the system's robust recognition of faces with ±50 degree range of full 3D head rotation.

In [21], an approach to pose-invariant face recognition was proposed. Gaussian mixture models with different numbers of mixture components are employed to characterize human faces and model pose variation. The optimal number of mixture components for each person is automatically learned from training data by growing the mixture models. The proposed algorithm is tested on real data recorded in a meeting room. The experimental results indicate that the new method outperforms the standard Eigenface and Gaussian mixture model approaches. This algorithm achieves as much as 42% error reduction compared with the standard Eigenface approach on the same test data.

5.2.2.2 Virtual image synthesis-based algorithms

Providing multiple gallery images to the recognition system is a rational method for dealing with pose variations. However, there are not always enough images available for many practical applications. An alternative solution is to augment the gallery by generating virtual views from one single face image, i.e., synthesizing virtual views under the desired

pose conditions. These methods, which make use of prior class information, are the most successful and practical methods currently available.

A. 2D Methods

To synthesize new view images, a good solution would be to use information about the three-dimensional nature of a face. However, 3D model-based approaches are computationally expensive and not easy to develop.

A linear object classes method was proposed in [9] that uses 2D example views of prototype faces under different rotations to represent 3D face prior knowledge. Experiments suggest that among the techniques for expressing prior knowledge of faces, 2D example-based approaches should be considered alongside the more standard 3D modeling techniques.

The underlying assumption of the linear object classes method is that the 3D shape of an object (and 2D projections of 3D objects) can be represented by a linear combination of prototype objects. It follows that a rotated view of the object is a linear combination of the rotated views of the prototype objects. This idea is used to synthesize rotated views of face images from a single example view [8][9]. To implement this method, a correspondence between images of the input object and a reference object is established using the optical flow method. Correspondences between the reference image and other example images having the same pose are also computed. Finally, the correspondence field for the input image is linearly decomposed into the correspondence fields for the examples. Compared with the parallel deformation scheme in [22], this method reduces the need to compute the correspondence between images of different poses. On the other hand, parallel deformation is able to preserve some peculiarities of texture that are nonlinear and that could be “erased” by linear methods. The recognition results vary depending on the available number of real views

of the person to be recognized. Without view synthesis, the recognition rate ranges from 66.7% when only one real view is available to 98.7% when 15 real views are available. By using view synthesis, the recognition rate is improved to 70-82% compared to the one-view baseline rate of 66.7%.

B. 2D Image + 3D generic model-based methods

A human face is a surface lying intrinsically in the 3D space. Therefore, a 3D model is better for representing faces, especially for handling variations such as pose and illumination. Blanz *et al.* [5][6] proposed a method based on a 3D morphable face model that encodes shape and texture in terms of model parameters, and an algorithm that recovers these parameters from a single image of a face. For face recognition, they use the shape and texture parameters of the models that are separated from the imaging parameters, such as pose and illumination.

The morphable face model in [5][6] is based on a vector space representation of faces. The database of laser scans used in this study contains scans of 100 males and 100 females recorded with a *Cyberware*^{TM} 3030PS scanner. Scans are stored in cylindrical coordinates relative to a vertical axis. The coordinates and texture values of all the n vertices of the reference face (n=75,972) are concatenated to form shape and texture vectors.

During image synthesis, new projected positions of vertices of the 3D model are rendered, along with illumination and color. During the process of fitting the model to a novel image, not only the shape and texture coefficients are optimized, but also the following rendering parameters, which are concatenated into a vector: the head orientation angles, the head position in the image plane, the size, the color and intensity of the light sources, the color constant, and the gain and offset of colors.

In [23], an appearance-based method, eigen light-fields, was introduced for face recognition across pose. This algorithm operates by estimating the light-fields of the subject's head. First, generic training data are used to compute an eigen-space of head light-fields, similar to the construction of eigenfaces; light-fields are simply used rather than images. Given a collection of gallery and probe images, projection onto the eigen-space is performed by setting up a least-square problem and solving for the projection coefficients using an approach similar to that used for dealing with occlusions in the eigen-space approach. Matching is performed by comparing the probe and gallery eigen light-fields.

This algorithm has several advantages. Any number of images can be used in both the gallery and probe sets. Moreover, none of the gallery images need to have been captured from the same pose as any of the probe images. For example, there might be two probe images for each person, a full frontal view and a full profile, and only one gallery image, a half profile. If only one probe or gallery image is available, the algorithm behaves "reasonably" when estimating the light-fields. If more than one probe or gallery image is available, the extra information (including the implicit shape information) is incorporated into a better estimate of the light-fields. The final face recognition algorithm therefore performs better with more input images.

In [24], a unified approach was proposed for solving both the pose and illumination problems. The basic idea is to use a varying-albedo reflectance model to synthesize new images in different poses from a real image. Using the generic 3D model, the authors approximately solved the correspondence problem involved in a 3D rotation, and performed an input-to-prototype image computation. To address the varying albedo issue in the estimation of both pose and light source, the use of a self-ratio image was proposed.

Improved recognition results are reported on a small database consisting of frontal and quasiprofile images of 115 novel faces.

To improve face recognition under pose variation, [25] presented a geometry-assisted probabilistic approach. The authors approximated a human head with a 3D ellipsoid model, so that any face image is a 2D projection of such a 3D ellipsoid at a certain pose. In this approach, both training and test images are back-projected to the surface of the 3D ellipsoid, according to their estimated poses, to form the texture maps. Thus, recognition can be accomplished by comparing the texture maps instead of the original images, as is done in traditional face recognition. In addition, the texture map is represented as an array of local patches, which enables the training of a probabilistic model for comparing corresponding patches. By conducting experiments on the CMU-PIE database, the proposed algorithm is shown to provide better performance than existing algorithms.

An analysis-by-synthesis framework for face recognition with variant pose, illumination and expression (PIE) was proposed in [26]. First, an efficient 2D-to-3D integrated face reconstruction approach is introduced to reconstruct a personalized 3D face model from a single frontal face image with neutral expression and normal illumination. Then, realistic virtual faces with different PIE are synthesized from the personalized 3D face to characterize the face subspace. Finally, face recognition is conducted based on these representative virtual faces. Compared with other related work, this framework has the following advantages: (1) only one single frontal face is required for face recognition, which avoids the burdensome enrollment work, (2) the synthesized face samples provide the capability to conduct recognition under difficult conditions, such as complex PIE, and (3) the proposed 2D-to-3D integrated face reconstruction approach is fully automatic and more efficient. The extensive

experimental results show that the synthesized virtual faces significantly improve the accuracy of face recognition with varying PIE.

In [27], a novel, pose-invariant face recognition system based on a deformable, generic 3D face model was presented. The model is a composite of: (1) an edge model, (2) a color region model, and (3) a wireframe model that jointly describes the shape and important features of the face. The first two submodels are used for image analysis and the third is mainly used for face synthesis. In order to match the model to face images in arbitrary poses, the 3D model is projected onto different 2D view planes based on rotation, translation and scale parameters, thereby generating multiple face-image templates of different sizes and orientations. Variations in face shape are taken into account by the deformation parameters of the model. Given an unknown face, its pose is estimated by model matching and the system synthesizes face images of known subjects in the same pose. The face is then classified as the subject whose synthesized image is most similar. The synthesized images are generated using a 3D face representation scheme that encodes the 3D shape and texture characteristics of the faces. This face representation is automatically derived from training face images of the subject. Experimental results show that this method is capable of determining the pose and recognizing a face accurately over a wide range of poses and with naturally varying lighting conditions. Recognition rates of 92.3% have been achieved by this method, with 10 training face images per person.

Table 5-2. Face recognition methods using multiview images.

Method	Database	Image Size	Subject No.	Image No. per Subject	Image Type	Training Image No. per Subject	Recognition Rate (%)
Real multiview [14]	Self-defined	n/a	62	25	Gray-level 2D	15	98%
View-based and modular Eigenface [15]	Self-defined	n/a	~3000	~2	Gray-level 2D	n/a	95%
AAMs for pose compensation [17]	CMU-PIE, IMM, TNO	640×486 n/a	68 54 189	n/a	Color 2D 2D 2D	3 3 3	75%
Illumination cone [18][19]	Yale B	640×480	10	576	Gray-level 2D	7	~100%
Parametric linear subspaces [20]	ATR	n/a	20	3625	Gray-level 2D	2821	98.7%
Growing Gaussian mixture models [21]	Self-defined	18×24 – 38×50	7	700	Gray-level 2D	500	42% error reduction
Linear object classes [22]	Self-defined	N/A	62	25	Gray-level 2D	1-15	82.2-98.7%
3D morphable model [6]	CMU-PIE, FERET	640×486 256×384	68 194	66 11	Color 2D Gray-level 2D	n/a	95% 95.9%
Eigen light-field [23]	CMU-PIE, FERET	640×486 256×384	68 200	13 9	Color 2D 2D	1 1	66.3% 75%
SFS [24]	FERET Stirling	Resized to 48×42	Select 108 pairs	n/a	Gray-level 2D	n/a	66.7%
Geometry assisted probabilistic modeling [25]	CMU-PIE	640×486	68	9	Color 2D	1	60-90+%
3D reconstruction [26]	CMU-PIE	640×486	68	~608	Color 2D	1	10-90+%
3D deformable model [27]	Self-defined	130×120	15	n/a	Color 2D	10	92.3%

5.2.3 3D Range Image-based Approach

As mentioned above, the majority of face recognition research and most commercial face recognition systems currently use normal 2D grayscale intensity images of the face. The 2D image + 3D generic model-based methods discussed in Section 5.2.2.2 only use a 3D model as an intermediate step in matching 2D images. In this sub-section, we discuss an approach that uses 3D range image data in face recognition [28]-[30]. Representative methods belonging to this class of face recognition system are summarized in Table 5-3.

A "range image", also sometimes called a "depth image," is an image where the pixel value reflects the distance from the sensor to the imaged surface [29]. The range image data can represent the 3D shape explicitly and can compensate for the lack of depth information in a 2D image. Therefore, it is argued in some papers, e.g., [30]-[33], that the 3D range data-based approach is more robust in dealing with pose variations. However, at the same time, the difficulty in obtaining 3D range data currently limits its wide applications.

3D face recognition research has been taking place since the 1980s. In [31], Cartoux *et al.* proposed a 3D face recognition method that segments a range image based on its principal curvature and finds a plane of bilateral symmetry through the face. This plane is used to normalize the pose. Partially because the database used in [31] is small (only 5 persons and 18 images in its dataset), a 100% rank-one recognition rate is reported in [31].

In [32], Lee *et al.* proposed a 3D face recognition method. This method segments the convex regions in the range image, based on the sign of the mean and the Gaussian curvatures. For each convex region, an extended Gaussian image (EGI) is created. A match between a region in a probe image and a region in a gallery image is found by cross-

correlating the EGIs. A graph-matching algorithm that incorporates the relational constraints is used to establish an overall match of the probe image with the gallery image.

In the 3D face recognition method proposed by Medioni *et al.* [33], an iterative closest point (ICP) approach was used to match face surfaces. Shapes are acquired by a passive stereo sensor. Experiments with 7 images each from a set of 100 subjects are reported, with the 7 images sampling different poses. An equal error rate (EER) of better than 2% is achieved.

In the 3D face recognition method studied in [34], Moreno *et al.* performed a segmentation based on Gaussian curvature and then create a feature vector based on the segmented regions. They examined the method and obtained the results on a dataset of 420 face meshes representing 60 different persons, with some sampling of different expressions and poses for each person. Recognition of 78% is achieved on the subset of frontal views.

Lee *et al.* [35] performed 3D face recognition by locating the nose tip, and then forming a feature vector based on contours along the face at a sequence of depth values. They reported 94% correct recognition at rank five.

In [36], Lu *et al.* also used an iterative closest point (ICP) based approach to 3D face recognition. This approach assumes that the gallery 3D image is a more complete face model and that the probe 3D image is a frontal view that is likely a subset of the gallery image. The database used includes images from 18 persons, with multiple probe images per person, incorporating some variations in pose and expression. Experiments with the database produced a recognition rate of 97%.

Utilizing the pose-invariant features of 3D face data, multiview face matching can potentially be accomplished. In [37], a feature extractor based on the directional maximum

was proposed to estimate the location of the nose tip and the pose angle simultaneously. A nose profile model represented by subspaces is used to select the best candidates for the nose tip. Assisted by a statistical feature location model, a multimodal scheme is presented to extract the corners of the eye and mouth. Using the automatic feature extractor, a fully automatic 3D face recognition system is developed. The system is evaluated on two databases, the MSU database (300 multiview test scans from 100 subjects) and the UND database (953 near frontal scans from 277 subjects). The automatic system provides recognition accuracy that is comparable to the accuracy of a system with manually labeled feature points.

There are other papers that use 3D techniques, e.g., [38]-[41], but as they mainly discuss face recognition under variation in expression, they are not discussed in this chapter.

Table 5-3. Face recognition methods using 3D range images.

Method	Database	Image Size	Subject No.	Image No. per Subject	Image Type	Training Image No. per Subject	Recognition Rate (%)
Profile extraction from range images [31]	Self-defined	n/a	5	18/5	3D	n/a	100%
[32]	Self-defined	256×150	6	1	3D	n/a	n/a
[33]	Self-defined	n/a	100	7	3D		98%
3D surface-extracted descriptors [34]	Self-defined	2.2K points	60	7	3D	n/a	78%
Local depth information [35]	Self-defined	320×320	35	2	3D	n/a	94% at rank 5
Matching 2.5D scans [36]	MSU	240×320	18	~6	3D	n/a	96%
Multiview 3D [37]	MSU, UND	n/a	100 277	3 ~3	3D 3D	n/a	96% 97%

5.3 Challenges to Pose-invariant Face Recognition

The invariant feature extraction-based algorithms can work well when the variations are not large. However, when the face images are highly varied due to severe pose changes, the sole dependency on feature extraction is still not enough.

Geometric model-based algorithms present promising performance for the pose-invariant face recognition task. However, these algorithms are still suffering from the difficulties and expensive computation of model generation, model fitting, and correspondence establishment.

It is often thought that systems using 3D facial images have potentially greater recognition accuracy than those using 2D images. 3D range images overcome the limitations of 2D images due to viewpoint and lighting variations, and therefore have the advantage of capturing shape variation irrespective of illumination variability. However, the difficulty in obtaining 3D range data currently limits its wide application.

Integrating 2D and 3D sensory information will be a key step in achieving a significant improvement in performance over systems that rely solely on a single type of sensory data. However, the multi-modal combination needs more sophisticated synergies between the two modalities in the interpretation of the data.

5.4 Conclusions

A review of typical face recognition algorithms in relation to one of the main obstacles, the pose variation problem has been presented in this chapter. A large number of face recognition algorithms, along with their modifications, have been developed during over the past decades. These algorithms have been categorized and briefly described. Future research challenges to pose-invariant face recognition have been identified. Although much effort has

been made to achieve pose-invariant face recognition and much progress has been achieved, the performance of such systems under pose variations still needs to be improved.

5.5 References

[1] L. Wiskott, J. Fellous, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 775-779, 1997.

[2] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681-685, 2001.

[3] P. Bellhumer, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Special Issue on Face Recognition, vol. 17, no. 7, pp. 711-720, 1997.

[4] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. on Computers*, vol. 42, pp. 300-311, 1993.

[5] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," *Proc. of SIGGRAPH'99*, pp. 187-194, 1999.

[6] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063-1074, 2003.

[7] V. Blanz, P. Grother, P. Phillips, and T. Vetter, "Face recognition based on frontal views generated from non-frontal Images," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 454-461, 2005.

[8] T. Vetter, "Learning novel views to a single face image," *Proc. of International Conference on Automatic Face and Gesture Recognition*, pp. 22-27, 1996.

[9] T. Vetter and T. Poggio, "Linear object classes and image synthesis from a single example image," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 733-742, 1997.

[10] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, 1991.

[11] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103-108, 1990.

[12] A. Lanitis, C. Taylor, and T. Cootes, "Automatic face identification system using flexible appearance models," *Image and Vision Computing*, vol. 13, pp. 393-401, 1995.

[13] B. Gokberk, L. Akarun, and E. Alpaydin, "Feature selection for pose invariant face recognition," *Proc. of International Conference on Pattern Recognition*, vol. 4, pp. 306-309, 2002.

[14] D. Beymer, "Face recognition under varying pose," *A. I. Memo No. 1461*, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1993.

[15] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 84-91, 1994.

[16] T. Cootes, G. Wheeler, K. Walker, and C. Taylor, "View-based active appearance models," *Image and Vision Computing*, vol. 20, pp. 657-664, 2002.

[17] P. Huisman, R. van Munster, S. Moro-Ellenberger, R. Veldhuis, and A. Bazen, **"**Making 2D face recognition more robust using AAMs for pose compensation," *Proc. of International Conference on Automatic Face and Gesture Recognition*, pp. 108-113, 2006.

[18] A. Georghiades, D. Kriegman, and P. Belhumeur, "Illumination cones for recognition under variable lighting: faces," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 52-58, 1998.

[19] A. Georghiades, P. Belhumeur and D. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, 2001.

[20] K. Okada and C. von der Malsburg, "Pose-invariant face recognition with parametric linear subspaces," *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition,* pp. 64-69, 2002.

[21] R. Gross, J. Yang, and A. Waibel, "Growing Gaussian mixture models for pose invariant face recognition," *Proc. of International Conference on Pattern Recognition*, vol. 1, pp. 1088-1091, 2000.

[22] D. Beymer and T. Poggio, "Face recognition from one example view," *Proc. of International Conference on Computer Vision*, pp. 500-507, 1995.

[23] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 4, pp. 449-465, 2004.

[24] W. Zhao and R. Chellappa, "SFS based view synthesis for robust face recognition," *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 285-292, 2000.

[25] X. Liu and T. Chen, "Pose-robust face recognition using geometry assisted probabilistic modeling," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 502-509, 2005.

[26] Y. Hu, D. Jiang, S. Yan, L. Zhang, and H. zhang, "Automatic 3D reconstruction for face recognition," *Proc. of IEEE International Conference on Automatic Face and Gesture*

Recognition, pp. 843-848, 2004.

[27] M. Lee and S. Ranganath, "Pose-invariant face recognition using a 3D deformable model," *Pattern Recognition*, vol. 36, pp. 1835-1846, 2003.

[28] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches to three-dimensional face recognition," *Proc. of International Conference on Pattern Recognition*, 2004.

[29] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multimodal 3D + 2D face recognition," *Computer Vision and Image Understanding*, vol. 101, pp. 1-15, 2006.

[30] A. Scheenstra, A. Ruifrok, and R. Veltkamp, "A survey of 3D face recognition methods," *Proc. of International Conference on Audio and Video Based Biometric Person Authentication*, pp. 891-899, 2005.

[31] J. Cartoux, J. Lapreste, and M. Richetin, "Face authentication or recognition by profile extraction from range images," *Proc. of Workshop on Interpretation of 3D Scenes*, pp. 194-199, 1989.

[32] J. Lee and E. Milios, "Matching range images of human faces," *Proc. of International Conference on Computer Vision*, pp. 722-726, 1990.

[33] G. Medioni and R. Waupolitsch, "Face recognition and modeling in 3D," *Proc. of IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 232-233, 2003.

[34] A. Moreno, A. Sanchez, J. Velez, and F. Diaz, "Face recognition using 3D surface-extracted descriptors," *Proc. of Irish Machine Vision and Image Processing Conference*, 2003.

[35] Y. Lee, K. Park, J. Shim, and T. Yi, "3D face recognition using statistical multiple features for the local depth information," *Proc. of International Conference on Vision Interface*, 2003.

[36] X. Lu, D. Colbry, and A. Jain, "Matching 2.5D scans for face recognition," *Proc. of International Conference on Pattern Recognition*, pp. 362-366, 2004.

[37] X. Lu and A. Jain, "Automatic feature extraction for multiview 3D face recognition," *Proc. of International Conference on Automatic Face and Gesture Recognition*, pp. 585-590, 2006.

[38] C. Chua, F. Han, and Y. Ho, "3D human face recognition using point signature," *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 233-238, 2000.

[39] A. Bronstein, M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," *International Journal of Computer Vision*, pp. 5-30, 2005.

[40] X. Lu and A. Jain, "Deformation analysis for 3D face matching," *Proc. of IEEE Workshop on Applications of Computer Vision*, pp. 99-104, 2005.

[41] K. Chang, K. Bowyer, and P. Flynn, "Adaptive rigid multi-region selection for handling expression variation in 3D face recognition," *Proc. of IEEE Workshop on Face Recognition Grand Challenge Experiments*, June 2005.

CHAPTER 6 FACIAL-COMPONENT-WISE POSE NORMALIZATION FOR POSE-INVARIANT FACE RECOGNITION[5]

6.1 Introduction

A practical face recognition system needs to work under varying imaging conditions, such as different pose, expression, and illumination conditions. In this chapter, we address one of these major issues, the pose variations.

It is well known that the performance of a face recognition system drops drastically when pose variations are present within the input images, and it has become a major goal to design algorithms that are able to cope with this kind of variations. While it is not difficult for human beings to recognize the same individual in varying poses, for automated systems, this is a difficult task. This is because the differences between two images of two different poses of a person would be more significant than the differences between two distinct persons in the same pose.

When a face is rotated in the image plane, it can be easily normalized by detecting at least two facial features. However, when the face is subjected to in-depth 3D rotation, simple geometrical image normalization is not possible. Many approaches have been proposed for automated face recognition under 3D rotation. Up to now, the most successful and practical algorithms are those that make use of prior knowledge of the class of faces. Having multi-view images of the same person stored in the gallery is one strategy for dealing with the pose-variant problem, and is a direct extension of frontal face recognition. An algorithm of this type was presented in [1]. In [2], the popular eigenface approach was extended to handle

[5] A version of this chapter has been submitted for publication, Shan Du and Rabab Ward, "Facial-Component-Wise Pose Normalization for Pose-Invariant Face Recognition."

multiple views. The authors compared the performance of a parametric eigenspace (computed using all views from all subjects) with view-based eigenspaces (separate eigenspaces for each view). In the experiments, the view-based eigenspaces outperform the parametric eigenspace. In [3], separate active appearance models [4] were trained for profile, half profile and frontal views.

Another popular solution is to generate virtual views. A generic 3D model of the human face can be used to predict the appearance of a face under different pose parameters [5]-[7]. Once a 2D face image is texture mapped onto the 3D model, the face can be treated as a traditional 3D object in computer graphics, undergoing 3D rotations. In [6], Blanz and Vetter proposed a 3D morphable model, where each face can be represented as a linear combination of 3D face exemplars. Given an input image, the 3D morphable model is fitted, recovering shape and texture parameters following an analysis-by-synthesis scheme. Blanz *et al.* also used the 3D morphable model in [7] to synthesize frontal faces from non-frontal views, which are then fed into the recognition system. In 3D methods, a precise 3D model of the face must be constructed from the current image. This requires the use of many techniques such as active camera calibration, feature point selection/detection, correspondent points labeling from different views, 3D model translation, rotation and projection, and a database of 3D heads. The enormous computational complexity involved may preclude such methods from becoming popular in real-world applications.

While the traditional approach relies on the use of 3D models, there are also some 2D example-based view synthesis methods that can generate virtual views under multiple poses. These methods exploit image transformations that are specific to the relevant object class, and are learnable from example views of other "prototype" objects of the same class. [8][9]

proposed an algorithm to synthesize novel views from a single image by using prior knowledge of facial images and apply them to face recognition. In this method, the prior face knowledge is represented by 2D views of prototype faces. The underlying assumption of this method is that the 3D shape of an object (and the 2D projections of 3D objects) can be represented by a linear combination of prototype objects. It follows that a rotated view of the object is a linear combination of the rotated views of the prototype objects. The so-called linear object classes (LOC) method makes use of this assumption to synthesize rotated views of facial images from a single sample view. In LOC, a facial image is first separated into a shape vector and a texture vector, and then LOC is applied to them respectively. The virtual "rotated" images are then easily generated using a base set of 2D prototype views. The synthesized virtual views are highly dependent on the correspondence between the prototype images. However, building accurate pixel-wise correspondence between facial images is a difficult problem.

In [10], a local linear regression (LLR) method, which starts from the basic idea of LOC, was proposed. The authors showed that, in the case where the given samples are well aligned, there exists an approximate linear mapping between two images of one person captured under variable poses. This mapping is consistent for all persons, if their facial images are aligned in a pixel-wise manner. Unfortunately, pixel-wise correspondence between images is still a challenging problem. In most real-world face recognition systems, facial images are only coarsely aligned, based on very few facial landmarks, such as the two eye centers. In this case, the above-mentioned assumption of linear mapping no longer holds theoretically, since it becomes a complicated nonlinear mapping. LLR proposes that by partitioning the

whole surface of the face into multiple uniform blocks, the linearity of the mapping for each block is increased because of a consistent normal and better control over alignment.

In this chapter, we start from the basic idea of LOC and LLR, that is, we try to generate the frontal view of a given non-frontal face image based on the 2D prototypes in a training set with corresponding image pairs of some specific poses. However, unlike previous methods [8]-[11], we apply the generation algorithm on multiple facial components rather than on separate shape and texture vectors (LOC) or on uniform image blocks (LLR). Accurate dense correspondence between face images is not required; what we need is just a coarse alignment based on the two eye centers. In our proposed method, the whole non-frontal face region is partitioned into multiple facial components where different normalization parameters are applied to different components for the generation of their frontal counterpart.

The remainder of this chapter is organized as follows: Section 6.2 describes the framework of the proposed method. In Section 6.3, the pose alignment algorithm is discussed in detail. The performance evaluation of the proposed method is presented in Section 6.4. Section 6.5 concludes the chapter.

6.2 Framework of the Proposed Method

We focus our attention on developing a technique for synthesizing face images that have different viewing positions from that of the given image, using only 2D views and information derived from prototype faces. The motivation behind using the example-based approach lies in its potential as a simple alternative to the more complicated 3D model-based approach.

The idea of segmenting an image into patches was inspired by LLR. For the case of coarse alignment, LLR shows that a local face region of the frontal view and its corresponding region in the non-frontal view satisfy the linearity assumption much better than the whole face regions (GLR). From the viewpoint of LOC, we can understand that estimating the linear combination coefficients will be easier and more accurate using small patches rather than the whole image.

Unlike LLR, which segments an image into uniform blocks, we partition the images according to the facial components positions. The reason for this innovation lies in our observation that using uniform blocks may break a facial component into pieces, i.e., the facial component may belong to more than one block. Moreover, the size of the blocks is not easy to select. If they are too large, the resulting image is blurred; and if they are too small, the coarse cross-pose correspondence may be meaningless, resulting in many annoying artifacts.

For our method, the selection of the patches according to facial component positions is more meaningful than in LLR, and therefore the establishment of coarse cross-pose correspondence is easier. The sizes of patches are neither too large nor too small. They are only related to the image size and are assigned automatically; no manual selection of the block size is needed, as in LLR. Also the patch size is not uniform - different components have different sizes. The facial components are also not broken into pieces. Thus, the blocking artifacts introduced by the block-based method will not ruin those facial components that are more important than others in face recognition.

The procedure for the proposed method is as follows (see Figure 6-1):

1. All images, including the training set and the probe image, are segmented into facial

components.

2. For each component patch,

   (a) the linear combination coefficients of the probe's non-frontal patch in terms of the training non-frontal patches (same pose) are computed.

   (b) the virtual frontal patch is generated using the above coefficients and the training frontal patches.

3. The virtual frontal patches are integrated to form the virtual frontal image.

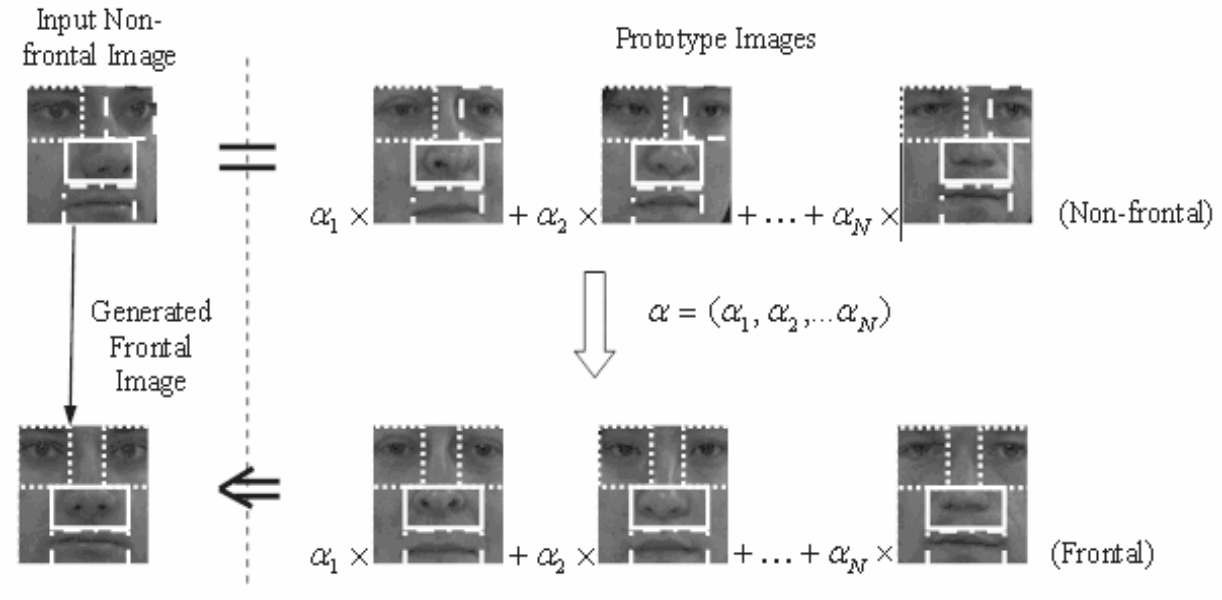


Figure 6-1. Component-based virtual frontal view generation.

6.3 Facial-Component-Wise Pose Normalization

Given a non-frontal face image, we generate its virtual frontal view based on a training set. We first represent the given non-frontal image using a linear combination of the training non-frontal images. Then, using these linear combination coefficients and the training frontal images, we generate the virtual frontal view of the given image. Because our method is component-based, we represent each component patch of the given non-frontal image as a

linear combination of the corresponding training non-frontal patches. Finally, using these linear combination coefficients and the corresponding training frontal patches, we generate the virtual frontal patch.

Estimating the linear combination coefficients for each patch becomes much easier and more accurate than estimating the global one because of the much lower dimension of the patches. This factor is especially important when the given training set is of limited size.

6.3.1 Component Segmentation

Since the positions of the two eyes' are already known, we can use them to roughly segment face images into facial components, as shown below.

It is well known from the art of drawing the human head that the average face is approximately five eye-lengths wide [12]. Both eyes lie on the line midway between the top of the face and the bottom of the chin, and the distance between them is approximately equal to one eye-length. The nose starts at the center of the face and descends to a point mid-way between the center of the face and the base of the chin. The width of the nose at the base is also equal to approximately one eye-length. The mouth barrel starts at the base of the nose and extends two thirds of the distance down from the nose to the chin. The distance between the central line of the mouth and the center of the face is approximately one third of the head length. The corners of the mouth align with the centers of the eye sockets.

Let (x_l, y_l) and (x_r, y_r) be the coordinates of the two eyes' centers. The distance between the two eyes is $d = x_r - x_l$. Please note that we do not consider the difference between y_l and y_r. Even though they are not equal, this difference is much smaller compared with the difference between x_l and x_r, and can therefore be ignored.

Normally, the distance between the inner corners of the two eyes is similar to the length of one eye. Therefore, we can use this information to segment a face (see Figure 6-2).

The face image is segmented into seven different-sized component patches. Each patch contains one facial component, e.g., eyes, the area between the eyes, nose, mouth, and cheeks.

As shown in Figure 6-2, moving from left to right and up to down, the sizes of the seven patches are calculated as follows:

1. Width: $\left(1 \to x_l + \frac{1}{4}d\right)$; Height: $\left(1 \to y + \frac{1}{4}d\right)$
2. Width: $\left(x_l + \frac{1}{4}d + 1 \to x_r - \frac{1}{4}d\right)$; Height: $\left(1 \to y + \frac{1}{4}d\right)$
3. Width: $\left(x_r - \frac{1}{4}d + 1 \to w\right)$; Height: $\left(1 \to y + \frac{1}{4}d\right)$
4. Width: $\left(1 \to x_l\right)$; Height: $\left(y + \frac{1}{4}d + 1 \to h\right)$
5. Width: $\left(x_l + 1 \to x_r\right)$; Height: $\left(y + \frac{1}{4}d + 1 \to y + \frac{1}{4}d + 1 + \frac{h - (y + \frac{1}{4}d)}{2}\right)$
6. Width: $\left(x_l + 1 \to x_r\right)$; Height: $\left(y + \frac{1}{4}d + 1 + \frac{h - (y + \frac{1}{4}d)}{2} + 1 \to h\right)$
7. Width: $\left(x_r + 1 \to w\right)$; Height: $\left(y + \frac{1}{4}d + 1 \to h\right)$

where w and h are the width and height of the image; $y = \max(y_l, y_r)$.

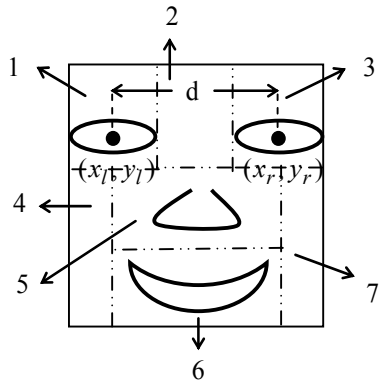


Figure 6-2. Component segmentation.

By using this segmentation, we can avoid breaking the facial features into pieces (i.e., each feature lies in one patch only). Moreover, since this segmentation directly results in meaningful component patches, the rough cross-pose correspondence is established automatically. Figure 6-3 shows the component segmentation on images with different poses.

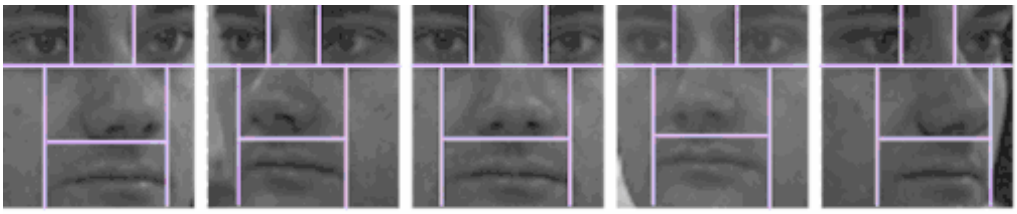


Figure 6-3. Component segmentation on images with different poses.

6.3.2 Coefficients Estimation

Given a non-frontal face image, our goal is to generate its virtual frontal view based on a training set. Simply speaking, we represent each component patch of the given non-frontal image using a linear combination of the corresponding training non-frontal patches. Then, using the linear combination coefficients and the corresponding training frontal patches, we generate the virtual frontal patch. The final virtual frontal image is generated by integrating the virtual frontal components. Estimating the linear combination coefficients for each patch

becomes much easier and more accurate than estimating the global one because of the much lower dimension of the patches. This factor is especially important when the given training set is of limited size.

Let $\left\{\left(\Phi^{p_0}, \Phi^{p_k}\right)\right\}$ be the training set of one component patch, where Φ^{p_0} denotes the frontal view p_0 composed of N subjects $\left\{x^{p_{0,1}}, x^{p_{0,2}}, ..., x^{p_{0,N}}\right\}$, and $\Phi^{p_k} = \left\{x^{p_{k,1}}, x^{p_{k,2}}, ..., x^{p_{k,N}}\right\}$ is the corresponding non-frontal view under pose p_k. Note that $x^{p_{k,i}}$ is the counterpart of $x^{p_{0,i}}$ from the same person but with different poses.

Following the LOC theory, we can use "prototype" 2D views and their known transformations to synthesize an operator that will transform a 2D view in pose p_k into a new 2D frontal view (pose p_0) when the object is a linear combination of the prototypes.

$$x^{p_k} = \sum_{i=1}^{N} \alpha_i x^{p_{k,i}} \tag{6-1}$$

$$x^{p_0} = \sum_{i=1}^{N} \alpha_i x^{p_{0,i}} \tag{6-2}$$

The decomposition of a given view x^{p_k} in (6-1) and the composition of the new view in (6-2) can be understood as a single linear transformation. First, we compute the coefficients α_i for the optimal decomposition (in the sense of least square). The given view is decomposed into the N "example" given prototypes by minimizing

$$\left\| x^{p_k} - \sum_{i=1}^{N} \alpha_i x^{p_{k,i}} \right\|^2 \tag{6-3}$$

We rewrite (6-3) as $x^{p_k} = \Phi^{p_k} \alpha$, where Φ^{p_k} is the matrix formed by the N vectors $x^{p_{k,i}}$ arranged column-wise, and α is the column vector of the α_i coefficients. Minimizing (6-3) gives

$$\alpha = (\Phi^{p_k})^{+} x^{p_k} \tag{6-4}$$

Then the new view x^{p_0} is given by

$$x^{p_0} = \Phi^{p_0}\alpha = \Phi^{p_0}\Phi^{p_k}{}^{+} x^{p_k} \tag{6-5}$$

and thus can be obtained from the 2D example pairs (Φ^{p_0}, Φ^{p_k}), where

$$\Phi^{p_k}{}^{+} = (\Phi^{p_k}{}^{T}\Phi^{p_k})^{-1}\Phi^{p_k}{}^{T} \tag{6-6}$$

6.3.3 Virtual Generation

The virtual frontal view can be obtained using Equation (6-5). After all virtual frontal components are generated; they are integrated to form the virtual frontal image.

6.4 Experimental Results

The proposed method is evaluated on the CMU-PIE face database. Both the visual quality of the generated virtual frontal views and the face recognition performance using the virtual images are presented. In the experiments, five pose subsets of the CMU-PIE database are used, which includes pose set 29 and 05 (turning left and right at 22.5 degrees), 11 and 37 (turning left and right at 45 degrees), and 27 (near frontal) [13]. The generation of the virtual frontal views use the leave-one-out strategy. In the final face recognition experiment, a total of 68 subjects are used with the frontal face images (pose 27) forming the gallery, while the non-frontal face images are used as probes to match against the frontal images in the gallery. The pose class and the face examples are given in Figure 6-4.

Figure 6-4. Face examples in CMU-PIE database.

In our experiments, face images are all normalized to the same size after fixing the eye positions and keeping the aspects of the faces. We implement four different recognition modes: without preprocessing (i.e., using the original non-frontal image directly as input), the global generation method (GLR) [10], the local generation method with uniform blocks (LLR) [10], and our proposed component-based generation method.

6.4.1 Virtual View Generation (Visual Quality)

In this section, some virtual view generation results are presented. The virtual frontal images are generated for all images from each of the 4 non-frontal pose sets. The virtual view generation can be decomposed into the input image reconstruction (Equation 6-1) and the virtual frontal view prediction (Equation 6-2). The results for the two stages are shown in Figure 6-5 and Figure 6-6, respectively.

In Figure 6-5, we show some examples of the input non-frontal image reconstruction results. From top to down, the pose sets are 29, 05, 11 and 37. Column (a) shows the input non-frontal images, column (b) the reconstructed non-frontal view generated by the global method GLR, columns (c) and (d) the results produced by the local method LLR with different block sizes, and column (e) the results generated by the component-based method.

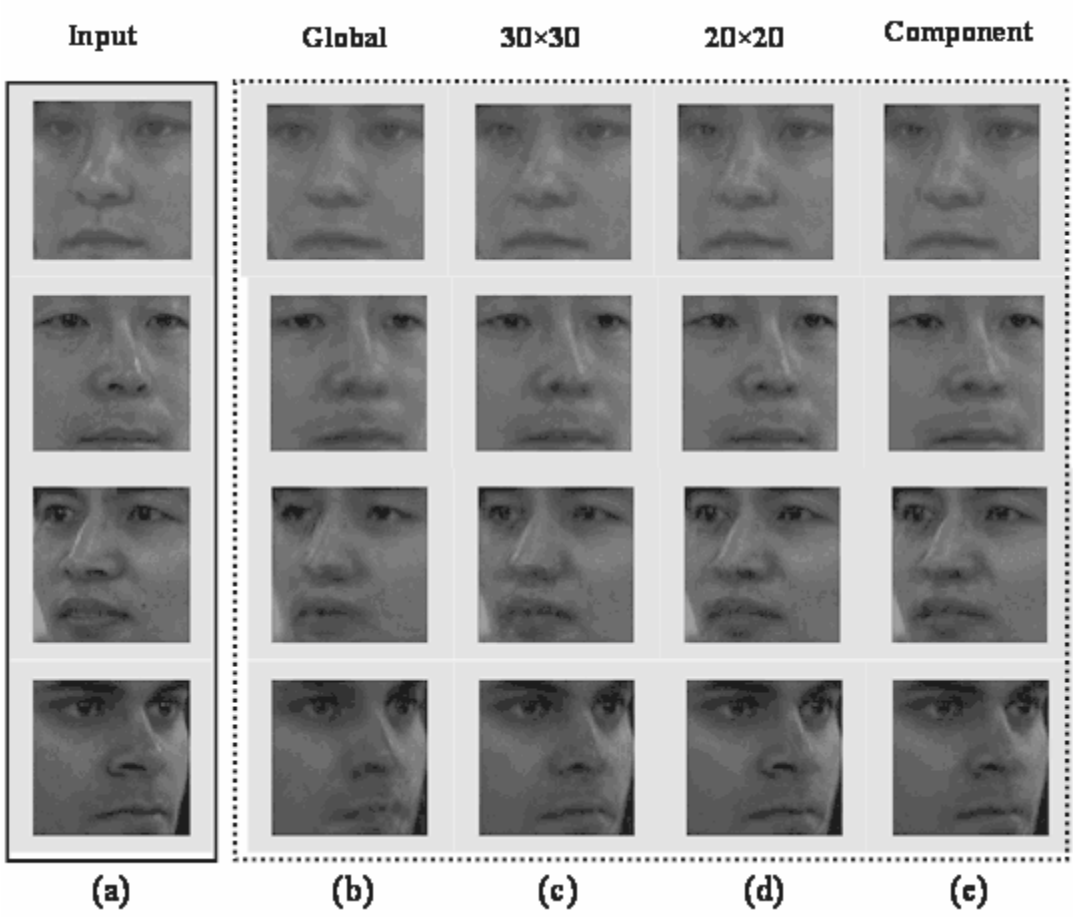


Figure 6-5. Examples of input non-frontal image reconstruction results.

In Figure 6-6, we show some examples of the virtual frontal view generation results. Column (a) shows the input non-frontal images, column (b) the virtual frontal view generated by the global method GLR, columns (c) and (d) the results produced by the local method LLR with different block sizes, and column (e) the results generated by the component-based method. The last column shows the real frontal faces. From these results, we can see that virtual generation using GLR is somewhat blurred; LLR can generate better results. In [10], the authors obtained the best results with block size 20×20. With the block size reduced to 10×10, the results became worse due to blocking artifacts. Compared with LLR, our method can generate a smoother image with fewer blocking artifacts. Most important, the facial

components are not broken into pieces. Figure 6-7 shows the virtual generated facial components. Compared with GLR and LLR, our method generates the best results. We do not break facial components into pieces. Thus, the blocking artifacts introduced by the patch-based method will not ruin those facial components that are more important than others in face recognition.

We have mentioned above that in LLR, the size of the blocks is manually selected. The authors obtained the best results with block size 20×20. However, if the image scale is changed, the best block size should be changed too. In contrast, our method assigns the patch sizes automatically regardless of the image scale, thus no manual selection of the block size is needed.

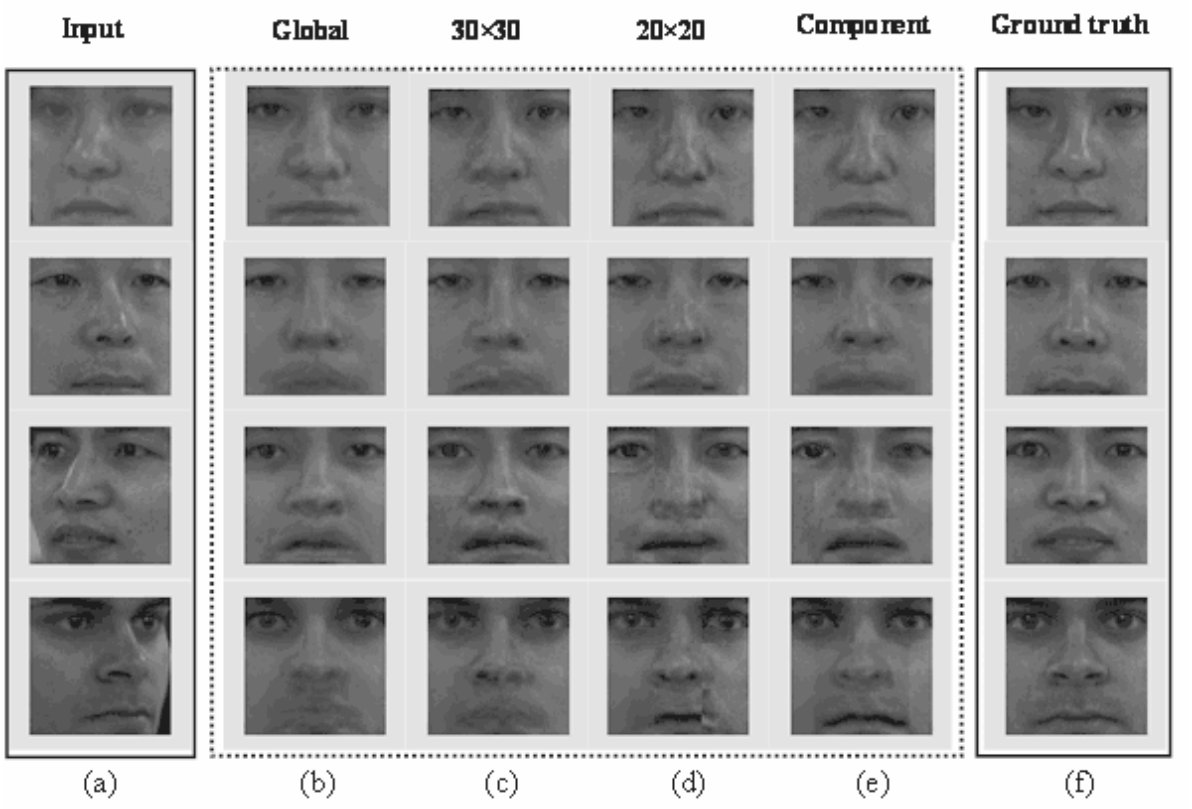


Figure 6-6. Examples of virtual frontal view generation results.

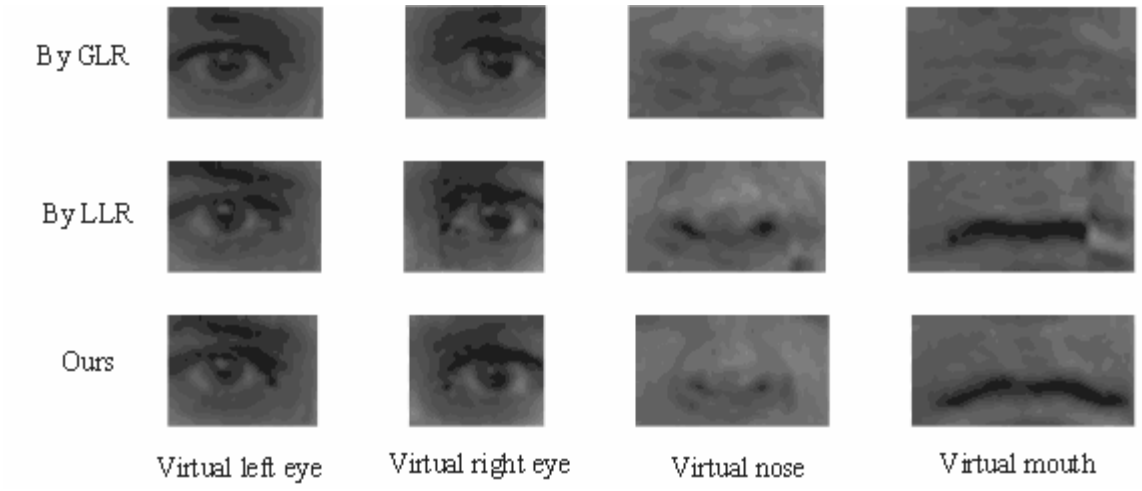


Figure 6-7. Virtual facial components.

In Figure 6-8, we show the virtual generation results on different scaled images. While the block size 20×20 can produce the best results on 60×60 image, it produces very bad results on larger images, e.g., 100×100 or 120×120 images. In contrast, our method always gets the best results irrespective of the image scale.

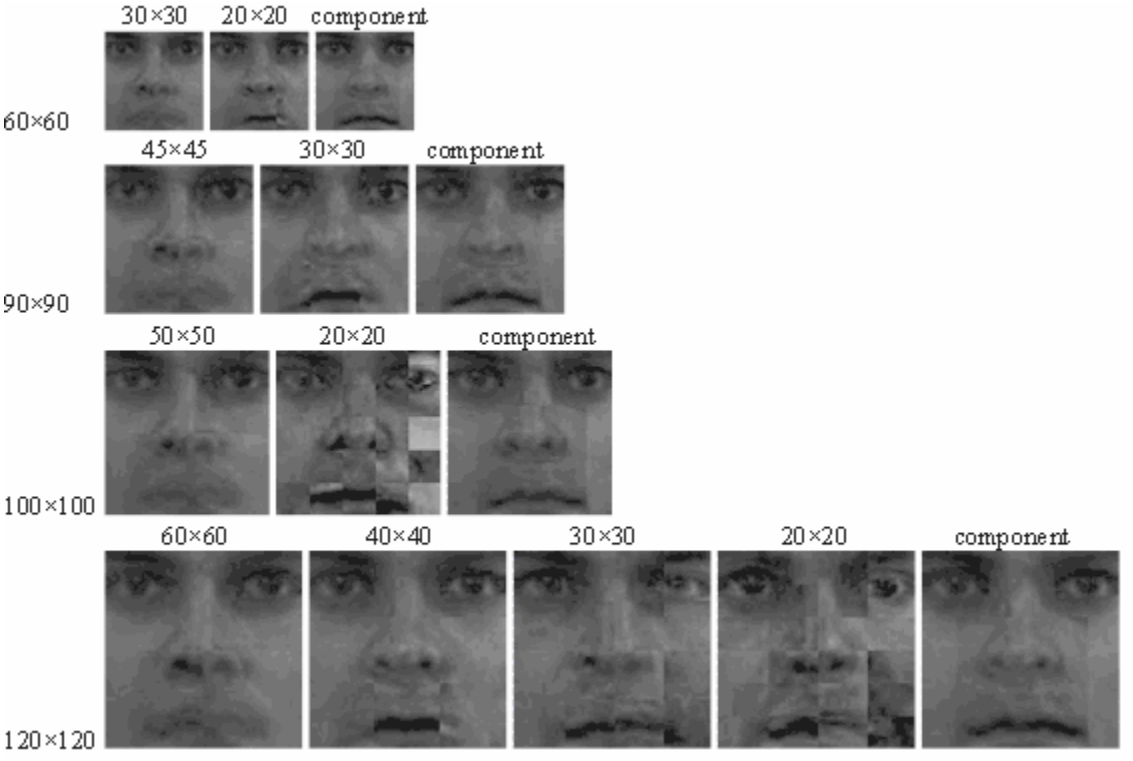


Figure 6-8. Virtual generation results on differently scaled images.

6.4.2 Peak Signal-to-Noise Ratio

To evaluate the virtual generation accuracy quantitatively, we compute the Peak Signal-to-Noise Ratio (PSNR) value of the generated image relative to its ground truth frontal image. The PSNR is calculated by

$$PSNR = 10 \times \log_{10} \frac{255 \times 255}{\frac{1}{wh} \sum_{i=1}^{w} \sum_{j=1}^{h} \left[I(i,j) - \hat{I}(i,j) \right]^2} \tag{6-7}$$

where $I(i,j)$ is the ground truth frontal image, $\hat{I}(i,j)$ is the generated virtual frontal image, and w and h are the width and height of the image, respectively.

Figure 6-9 shows the PSNR values of different generated images. Our proposed method generate the best image.

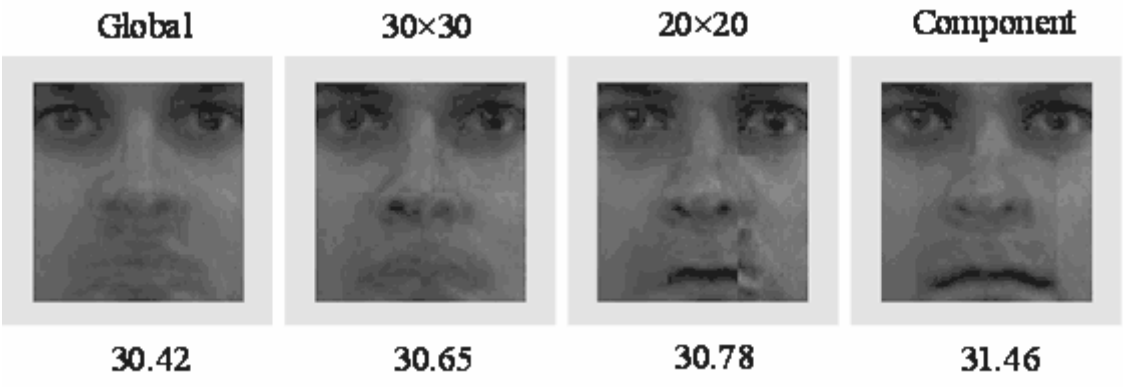


Figure 6-9. PSNR for virtual images.

6.4.3 Pose-invariant Face Recognition using Virtual Views

In this section, we describe the pose-invariant face recognition experiments that we carry out on the virtual frontal views to evaluate the proposed algorithm.

We implement four different recognition modes: without preprocessing (original), the global generation method (GLR) [10], the local generation method with uniform blocks

(LLR) [10] (with block sizes of 30×30 and 20×20), and our proposed component-based generation method.

From Figure 6-10, it can be seen that our method outperforms the others.

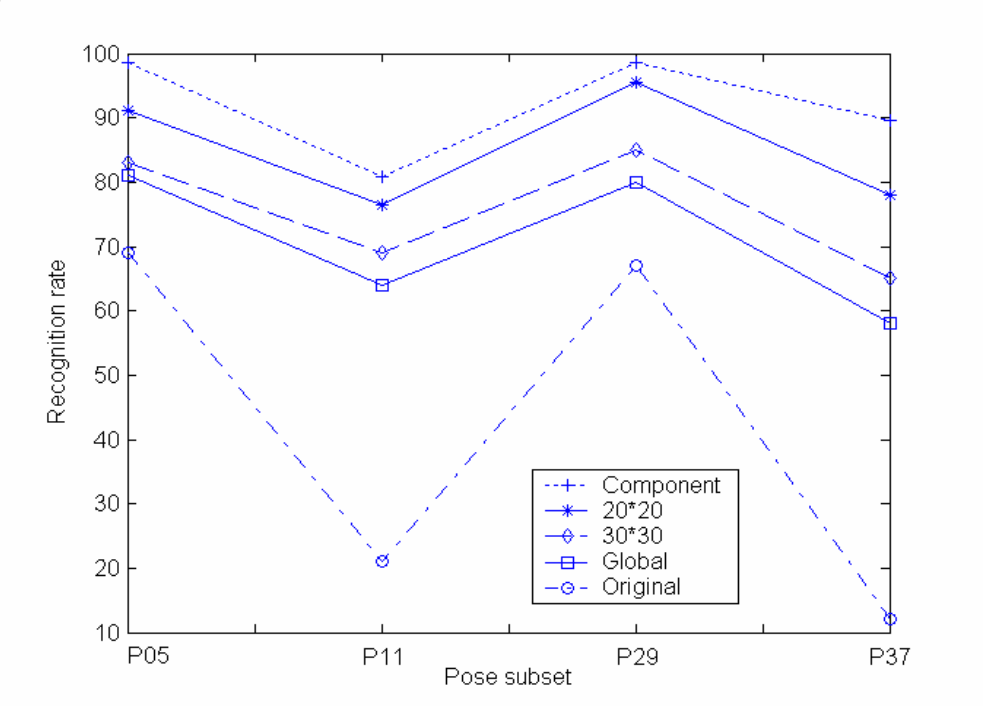


Figure 6-10. Recognition rate comparison.

In Table 6-1, we also show the comparison of our method with the Eigen Light-Field (ELF) method [15] that is well known for recognizing faces across pose and achieving good performance. The ELF method operates by estimating the light-fields of the subject's head. First, generic training data are used to compute an eigen-space of head light-fields, similar to the construction of eigenfaces; light-fields are simply used rather than images. Given a collection of gallery and probe images, projection onto the eigen-space is performed by setting up a least-square problem and solving for the projection coefficients using an approach similar to that used for dealing with occlusions in the eigen-space approach. Matching is performed by comparing the probe and gallery eigen light-fields. Our method

outperforms the ELF method.

Table 6-1. The performance comparison between our method and other methods.

Methods	P05	P11	P29	P37
ELF [15]	88%	76%	86%	74%
LLR (20×20) [10]	91.2%	76.5%	95.6%	77.9%
Our method	**98.5%**	**80.9%**	**98.5%**	**89.7%**

6.5 Conclusions

In this chapter, we propose an efficient facial-component-based pose normalization method for pose-invariant face recognition. The effectiveness of the proposed method is evaluated by the face recognition experiments on the CMU-PIE database. The experimental results show that partitioning face images into facial components is more meaningful than partitioning them into uniform blocks, resulting in better performance in terms of both visual quality and recognition rate.

6.6 References

[1] D. Beymer, "Face recognition under varying pose," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 756-761, 1994.

[2] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 84-91, 1994.

[3] T. Cootes, G. Wheeler, K. Walker, and C. Taylor, "View-based active appearance models," *Image and Vision Computing*, vol. 20, pp. 657-664, 2002.

[4] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681-685, 2001.

[5] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," *Proc. of SIGGRAPH'99*, pp. 187-194, 1999.

[6] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063-1074, 2003.

[7] V. Blanz, P. Grother, P. Phillips, and T. Vetter, "Face recognition based on frontal views generated from non-frontal images," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 454-461, 2005.

[8] T. Vetter, "Learning novel views to a single face image," *Proc. of International Conference on Automatic Face and Gesture Recognition*, pp. 22-27, 1996.

[9] T. Vetter and T. Poggio, "Linear object classes and image synthesis from a single example image," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 733-742, 1997.

[10] X. Chai, S. Shan, X. Chen, and W. Gao, "Local linear regression (LLR) for pose invariant face recognition," *Proc. of International Conference on Automatic Face and Gesture Recognition,* pp. 631- 636, 2006.

[11] C. Hsieh and Y. Chen, "Kernel-based pose invariant face recognition," *Proc. of IEEE International Conference on Multimedia & Expo*, pp. 987-990, 2007.

[12] B. Hogarth, *Drawing the Human Head.* 1st ed., New York: Watson-Guptill, 1965.

[13] T. Sim, S. Baker and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615-1618, 2003.

[14] P. Huisman, R. van Munster, S. Moro-Ellenberger, R. Veldhuis, and A. Bazen, "Making 2D face recognition more robust using AAMs for pose compensation," *Proc. of*

International Conference on Automatic Face and Gesture Recognition (FGR), pp. 108-113, 2006.

[15] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 4, pp. 449-465, 2004.

CHAPTER 7 SUMMARY

7.1 Major Thesis Contributions

The objective of this thesis is to investigate appropriate face recognition techniques that are robust against two significant variations involved in face images, pose and illumination variations. We have decided to tackle the problems and achieve the research objective from two aspects: (1) to propose a new face feature extraction and representation technique, and (2) to propose new problem-specific pre-processing techniques.

Large amount of experiments have shown that our proposals are effective and the objective of my doctoral research has been reached. The major contributions of my research and the thesis are summarized as follows.

The first contribution is **a new face feature extraction and representation method** that employs non-uniform multi-level selection of Gabor features.

The new method is based on the local statistics of the Gabor features and is implemented using a coarse-to-fine hierarchical strategy. Gabor features that correspond to important face regions are selected automatically and sampled finer than other features. The sampling rate adaptation is implemented automatically without facial landmark position initialization. The non-uniformly extracted Gabor features are then classified using principal component analysis and/or linear discriminant analysis for the purpose of face recognition. To verify the effectiveness of the proposed method, experiments are conducted on the ORL, Yale, Yale B, and FERET face image databases, where the images vary in illumination, expression, pose, and scale. Besides its advantage in yielding significantly higher recognition accuracy, the proposed method greatly reduces the dimensionality of the features for classification and is thus computationally less demanding. The experimental results also show that the proposed

method works well not only when multiple sample images are available for each person but also when only one sample image is available for training. By using the new non-uniform multi-level face representation method, the overall system performance is substantially improved. The new method has been discussed in detail in Chapter 2 of the thesis, published in part in IEEE ICASSP 2005 (best paper award) [1] and accepted for publication in the IEEE Transactions on Systems, Man and Cybernetic, Part B [2] .

Though the above feature extraction method yields good performance under some degree of pose and illumination variations, when the face images are highly varied due to pose and illumination changes, the sole use of feature extraction is still not enough. We strongly believe that adding specifically designed pre-processing methods before feature extraction can substantially increase face recognition performance in cases of heavy variations, which is another aspect of my work.

To solve the varying illumination problem, especially the side lighting effect problem in face recognition, we proposed **a novel adaptive region-based image preprocessing scheme** that enhances face images and facilitates the illumination invariant face recognition task. This constructs the second major contribution of this thesis.

The region-based method first segments an image into different regions according to its different local illumination conditions, then both the contrast and the edges are enhanced regionally so as to alleviate the side lighting effect. Since illumination variations mainly lie in the low-frequency band, the proposed contrast enhancement scheme uses adaptive region-based histogram equalization (ARHE) of the low-frequency coefficients to minimize variations under different lighting conditions. By observing that under poor illuminations the high-frequency features are more important in recognition, we enlarge the high-frequency

coefficients to make face images more distinguishable. Compared with existing image preprocessing methods, our method is shown to be more suitable for dealing with uneven illuminations in face images. Experimental results show that the proposed method significantly improves the performance of face images with illumination variations. The proposed method does not require any modeling or model fitting steps and can be implemented easily. Moreover, it can be applied directly to any single image without using any lighting assumption, and any prior information on 3D face geometry. The proposed method has been detailed in Chapter 3 of the thesis, published in part in IEEE ICIP 2005 and IEEE ICASSP 2006 [3][4] and submitted for publication in the IEEE Transactions on Circuits and Systems for Video Technology [5].

Accurate detection of eyes in face images can substantially improve face recognition performance. We have proposed **an efficient method that can automatically detect the positions of eyes** in gray-scale face images and is robust to illumination and pose variations. This method constitutes the third major contribution of the thesis.

The approach does not require any prior knowledge about face orientation and illumination strength. It does not need an initialization and training process which is often time and manpower consuming in existing methods. This approach consists of four steps. Based on an edge map obtained via multi-resolution wavelet transform, the approach first segments an image into different inhomogeneously illuminated regions. The illumination of every region is then adjusted separately so that the features' details are more pronounced. To locate the different facial features, for every region, a Gabor-based image is constructed from the illumination adjusted image. The eyes sub-regions are then identified using the edge map of this image. The regionally illumination adjusted gray-scale image, the regionally obtained

Gabor image, and the regionally obtained edge map are used altogether to detect the eyes. This method has been applied successfully to the images of the Yale B face database that have different illuminations and different poses. The experimental results show that this method works well for face images of different people and under various conditions such as changes in lighting and pose. The proposed method has been detailed in Chapter 4 of the thesis, published in part in IEEE ICIP 2007 [6] and submitted for publication in the IEEE Transactions on Information Forensics and Security [7].

The fourth major contribution is the introduction of **a novel facial-component-wise pose normalization method** for facilitating pose-invariant face recognition.

The main idea is to normalize a non-frontal facial image to a virtual frontal image component by component. In this method, we first partition the whole non-frontal facial image into different facial components and then the virtual frontal view for each component is estimated separately. The final virtual frontal image is generated by integrating the virtual frontal components. The proposed method relies only on 2D images, therefore complex 3D modeling is not needed. The effectiveness of the proposed method is evaluated by face recognition experiments on the CMU-PIE database. The experimental results show that partitioning facial images into facial components is more meaningful than partitioning them into uniform blocks, resulting in better pose normalization results in both visual quality and recognition rate. The experimental results demonstrate the advantages of the proposed method over the local linear regression (LLR) method and the eigen light-field (ELF) method. The proposed method has been presented in detail in Chapter 6 of the thesis, and submitted for publication in the IEEE Transactions on Multimedia [8].

In addition to the four major contributions, a complete survey of typical face recognition

algorithms has been conducted, particularly in relation to one of the main obstacles, the pose variation. This survey has been presented in detail in Chapter 5 of the thesis, and published in the Journal of the Franklin Institute [9].

7.2 Discussions on Future Work

As shown by numerous experiments and intensive study, the methods proposed in this thesis are effective and can substantially improve face recognition performance for many practical cases.

Following the research achievements of the thesis, it is worthwhile to integrate all the proposed methods into a practical engineering tool for face recognition. However, implementation of the tool requires a significantly large amount of software programming work and manpower. We thus leave this task for future work.

In order to become a prominent research group in face recognition, it is also suggested that the UBC image processing lab builds a face database with self proprietary. Some face databases are available around the world, e.g., FERET, CMU-PIE, Yale B, etc., but during my research, I found that obtaining appropriate face databases have often met some obstacles, such as, long delay of database delivery from the owners, and no access to some databases. Also restricted by existing databases, the research can not be freely extended to some potentially new topics that may emerge and produce prominent fruits. However, the accumulation of new face images and the establishment of the databases are very time consuming. This can be a long term future work.

The proposed new methods have been proven to be effective and efficient in improving the face recognition performance. However, there are still some interesting studies that can be followed as an extension of this thesis.

It is often thought that systems using 3D facial images have potentially greater recognition accuracy than those using 2D images. 3D range images overcome the limitations of 2D images due to viewpoint and lighting variations, and therefore have the advantage of capturing shape variation irrespective of illumination variability. However, the difficulty in obtaining 3D range data currently limits its wide application.

Integrating 2D and 3D sensory data will be a key step in achieving a significant improvement in performance over systems that rely solely on a single type of sensory data. However, this multi-modal combination needs more sophisticated synergies between the two modalities in the interpretation of the data. This needs a further intensive investigation.

7.3 References

[1] S. Du and R. Ward, "Statistical non-uniform sampling of Gabor wavelet coefficients for face recognition," *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* vol. 2, pp. 73-76, 2005.

[2] S. Du and R. Ward, "Improved face representation by non-uniform multi-level selection of Gabor convolution features," *IEEE Trans. on Systems, Man and Cybernetic, Part B*, accepted for publication.

[3] S. Du and R. Ward, "Wavelet-based illumination normalization for face recognition," *Proc. of IEEE International Conference on Image Processing (ICIP)*, vol. 2, pp. 954-957, 2005.

[4] S. Du and R. Ward, "Adaptive region-based image enhancement method for face recognition under varying illumination conditions," *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, pp. 353-356, 2006.

[5] S. Du and R. Ward, "Adaptive region-based image enhancement method for robust face recognition under variable illumination conditions," *IEEE Trans. on Circuits and Systems for Video Technology*, submitted.

[6] S. Du and R. Ward, "A robust approach for eye localization under variable illuminations," *Proc. of IEEE International Conference on Image Processing (ICIP)*, vol. 1, pp. 377-380, 2007.

[7] S. Du and R. Ward, "Eye detection in gray-scale face images under various illumination conditions," *IEEE Trans. on Information Forensics and Security*, submitted.

[8] S. Du and R. Ward, "Facial-component-wise pose normalization for pose-invariant face recognition," *IEEE Trans. on Multimedia*, submitted.

[9] S. Du and R. Ward, "Face recognition under pose variations," *Journal of the Franklin Institute*, vol. 343, no. 6, pp. 596-613, 2006.