Recognizing Faces --- An Approach Based on Gabor Wavelets

By LinLin Shen, BSc, MSc

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

July 2005

Abstract

As a hot research topic over the last 25 years, face recognition still seems to be a difficult and largely problem. Distortions caused by variations in illumination, expression and pose are the main challenges to be dealt with by researchers in this field. Efficient recognition algorithms, robust against such distortions, are the main motivations of this research.

Based on a detailed review on the background and wide applications of Gabor wavelet, this powerful and biologically driven mathematical tool is adopted to extract features for face recognition. The features contain important local frequency information and have been proven to be robust against commonly encountered distortions. To reduce the computation and memory cost caused by the large feature dimension, a novel boosting based algorithm is proposed and successfully applied to eliminate redundant features. The selected features are further enhanced by kernel subspace methods to handle the nonlinear face variations. The efficiency and robustness of the proposed algorithm is extensively tested using the ORL, FERET and BANCA databases.

To normalize the scale and orientation of face images, a generalized symmetry measure based algorithm is proposed for automatic eye location. Without the requirement of a training process, the method is simple, fast and fully tested using thousands of images from the BioID and BANCA databases.

An automatic user identification system, consisting of detection, recognition and user management modules, has been developed. The system can effectively detect faces from real video streams, identify them and retrieve corresponding user information from the application database. Different detection and recognition algorithms can also be easily integrated into the framework.

List of Publications

Some parts of the work presented in the thesis have been published in the following articles:

1. LinLin Shen and Li Bai. Information theory for Gabor feature selection for face recognition. *Eurasip Journal on Applied Signal Processing*, in press, 2005.

2. LinLin Shen and Li Bai. A review on Gabor wavelets for face recognition. Revision submitted, *Pattern Analysis and Application*, 2005.

3. LinLin Shen and Li Bai. A fast and robust Gabor feature based method for face recognition. In revision (invited submission), *Special Issue on Crime Detection and Prevention, Pattern Recognition Letters*, 2005.

4. LinLin Shen, Li Bai, Daniel Bradsley and YangSheng Wang. Gabor feature selection using improved AdaBoost learning. *The International Workshop on Biometric Recognition Systems, in conjunction with ICCV'05*, Beijing, 2005.

5. LinLin Shen and Li Bai. Kernel enhanced informative Gabor features for face recognition. *The 16th British Machine Vision Conference (BMVC)*, Oxford, 2005.

6. Li Bai and LinLin Shen. A fast and robust Gabor feature based method for face recognition. *The IEE International Symposium on Imaging for Crime Detection and Prevention*, The IEE Savoy Place, London, 2005.

7. LinLin Shen and Li Bai. AdaBoost Gabor feature selection for classification. *Proc. of Image and Vision Computing New Zealand*, New Zealand, 2004.

8. LinLin Shen, Li Bai and P., Picton. Facial recognition/verification using Gabor wavelets and kernel methods. *Proc. of the IEEE Internal Conference on Image Processing (ICIP)*, Singapore, 2004.

9. LinLin Shen and Li Bai. Combining Gabor feature and Kernel Direct Discriminant Analysis for face recognition. *Proc. of the 17th International Conference on Pattern Recognition (ICPR)*, Cambridge, UK, Aug., 2004.

10. Kieron Messer, Josef Kittler, Mohammad Sadeghi, Miroslav Hamouz, Alexey Kostin , Fabien Cardinaux, Sebastien Marcel, Samy Bengio, Conrad Sanderson, Norman Poh, Yann Rodriguez, Jacek Czyz, L. Vandendorpe, Chris McCool, Scott Lowther, Sridha Sridharan, Vinod Chandran, Roberto Parades Palacios, Enrique Vidal, Li Bai, LinLin Shen, Yan Wang, Chiang Yueh-Hsuan, Liu Hsien-Chang, Hung Yi-Ping, Alexander Heinrichs, Marco Müller, Andreas Tewes, Christoph von der Malsburg, Rolf Würtz, Zhenger Wang, Feng Xue, Yong Ma, Qiong Yang, Chi Fang, Xiaoqing Ding, Simon Lucey, Ralph Goss, and Henry Schneiderman. Face authentication test on the BANCA database. *Proc. of the 17th International Conference on Pattern Recognition (ICPR)*, Cambridge, UK, 2004.

11. LinLin Shen and Li Bai. Gabor feature based face recognition using Kernel methods. *Proc. of the IEEE 6th International Conference on Automatic Face and Gesture Recognition,* Soel, Korea, May, 2004.

12. Li Bai and LinLin Shen. Combining wavelets with HMM for face recognition. *Proc. of the 23rd SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, Cambridge, UK, Dec., 2003.

13. LinLin Shen and Li Bai. A comparison of eye and non-eye classifiers, *Proc. of the 1st Annual Academic Conference of CSSA-Nottingham On Science and Engineering*, Nottingham, UK, Jul., 2003, pp.30-40.

14. Li Bai and LinLin Shen. Face detection by orientation map matching, *Proc. of International Conference on Computational Intelligence for Modelling Control and Automation*, Austria, Feb., 2003, pp. 363-369.

15. LinLin Shen and Li Bai. Effects of different Gabor filter parameters on face recognition, *Technical Report*, School of CS & IT, University of Nottingham, 2004.

16. LinLin Shen and Li Bai. PCA versus LDA for face verification using small size training data, *Technical Report*, School of CS & IT, University of Nottingham, 2003.

17. LinLin Shen and Li Bai. A video based real time face recognition system, *Technical Report*, School of CS & IT, University of Nottingham, 2003.

Acknowledgement

I would like to thank my supervisor Dr. Bai Li for her valuable suggestions, encouragement and helps through the whole course of my research.

I would like to thank Professor Peter Ford for his consistent encouragement and support to our research.

Also thanks to Mr. Daniel for checking grammars of the thesis and the technical staff, Viktor, for his friendly and generous help during the three years.

I would like to express my deepest gratitude to my parents, sister and brothers. Without their support and love, I would not have studied anything.

This thesis is dedicated to my wife, Shandy, ZhengXiang. Without her support and encouragement, there would never be any chance for this thesis to happen.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

The major concern of this thesis is to develop an automatic face recognition system, which is robust against variance in illumination, expression and pose. At the same time, the system has to take computation and memory cost into consideration for real time applications. This chapter will give a brief introduction to the background of this research and a summary of some potential applications. Following the description on how to evaluate the performance of different systems, motivations behind the research and the organization of the thesis will be introduced.

1.1 Automatic Person Identification

With the advent of electronic banking, e-commerce, smartcards and an increased emphasis on the privacy and security of information stored in various databases, automatic personal identification has become a very important topic. Accurate automatic personal identification is now needed in a wide range of civilian applications involving the use of passports, cellular phones, automatic teller machines and driver license. Traditional knowledge-based (password or Personal Identification Number (PIN)) and token-based (passport, driver license, and ID card) identifications are prone to fraud because PINs may be forgotten or guessed by an impostor and the tokens may be lost or stolen. Therefore, traditional knowledge-based and token-based only approaches are unable to satisfy the security requirements of our electronically interconnected information society. A perfect identity authentication system will need a biometric component.

1.2 Biometrics

A biometric is a representation of a unique part or characteristic of an individual which has the potential capability to distinguish between an authorised person and an impostor. Since biometric characteristics are distinctive, cannot be forgotten or lost, and the person to be authenticated needs to be physically present at the point of identification, biometrics are inherently more reliable and more capable than traditional knowledge-based and token-based techniques. Currently there are many biometric technologies used for personal authentication: face, fingerprint, hand geometry, iris, retina, signature, voice, etc. Despite the fact that other methods of identification (such as fingerprint, or iris scans) can be more accurate, face recognition has always remained a major focus of research because of its non-invasive nature and because it is human's primary method of identification. The technology of face recognition can be widely applied in security

surveillance, authentication, access control and human computer interfaces. Since the late eighties there has been an explosive growth in research on face recognition because of the practical importance of the topic and theoretical interest from both cognitive scientists and computer vision and pattern recognition researchers.

1.3 Face Identification and Verification

A biometric system can be operated in two modes: verification mode and identification mode. In the verification mode, a biometric system either accepts or rejects a user's claimed identity while a biometric system operating in the identification mode establishes the identity of the user without a claimed identity. Face identification is a more difficult problem than face verification because a huge number of comparisons need to be performed in order to complete identification. There are a number of potential civilian applications for a biometric system working in verification mode. For example, an ATM system which verified a user's face with a biometric upon each transaction would need only to match the current face image (acquired at point of transaction) with a single template stored on the ATM card. A typical face verification system can be divided into two modules: enrolment and verification. The enrolment module scans the face of a person through a sensing device and then stores a representation (template) of the face in the database. The verification module is invoked during the operation phase. The same representation used in enrolment phase is extracted from the input face and matched against the template of the claimed identity to give a "yes/no" answer. On the other hand, an identification system matches the input face with a large number of faces in the database and as a result, algorithm efficiency is a critical issue in an identification system.

1.4 Performance Evaluation

1.4.1 Identification System

A face identification systems performance is usually evaluated by recognition rate, which is calculated by matching a set of test face images with those in the database. Different algorithms can be evaluated by matching each test face image. The matching attempts performed for each test usually consist of correct matches and incorrect matches. A matching is considered as correct if the two face images being matched are from the same person, and incorrect otherwise. Recognition rate is defined as the ratio between the number of correct matches and the number of test images.

1.4.2 Verification System

In a face verification system, system level performance evaluations are usually performed by cross matching the face images in the database. Different algorithms can be evaluated by matching each face image in the database with the rest of the images in the database. A threshold value is normally used such that a matching attempt is considered authentic when the matching score is equal or above the threshold value. Two metrics (FAR and FRR) are used to measure performance of the whole system. The false acceptance rate, or FAR, is the measure of the likelihood that the biometric security system will incorrectly accept an access attempt by an unauthorized user. A system's FAR typically is stated as the ratio of the number of false acceptances divided by the number of impostor attempts. The false rejection rate, or FRR, is the measure of the likelihood that the biometric security system will incorrectly reject an access attempt by an authorized user. Analysis of the FAR shows how well the system can distinguish a correct match from an incorrect match and is usually related to the uniqueness of the features. On the other hand, FRR analysis focuses on the repeatability

of the features between different faces of the same person and is related to the reliability of the features.

A system can be tuned for a particular application by varying the value of these two metrics. A low value for both metrics is often desirable. Unfortunately, trying to minimise *FAR* or *FRR* requires a trade off between each of the metrics. The Receiver Operating Curve (ROC) plots *FAR* versus *FRR* (Jonsson, Kittler, Li, & Matas, 2002) for a system and can be used as a guide for the selection of an operating point for the system. The *FAR* is usually plotted on the horizontal axis as the independent variable and the *FRR* is plotted on the vertical axis as the dependent variable. The closer the ROC-curve to the x and y axes, the lower verification error and thus the more reliable the system. In reporting the performance, the values of *FAR* and *FRR* for the ROC-curve are computed by varying the threshold value and using:

$$FAR = \frac{n_{ac}}{n_u}; \quad FRR = \frac{n_{re}}{n_a} \tag{1.1}$$

In Equation (1.1), n_a is the number of access attempt by an authorized user and n_u is the number of access attempt by an unauthorized user. For a given threshold value, n_{ac} is the number of acceptances and n_{re} is the number of rejections. From the ROC-curve, the Equal Error Rate (*EER*) is defined as the point where the value of *FAR* equals the value of *FRR*. The value of *EER* can now be used to determine the performance of the system. The lower is the value of *EER*, the more reliable the system.

1.5 Motivation and Solutions

As a hot research topic over the last 25 years, a large number of face recognition algorithms have been proposed in the literature. The next chapter contains a detailed survey of this research. With a number of different databases available, it is always very difficult to compare different face recognition algorithms. Even when the same database

is used, researchers may use different protocols for testing. Whilst many of the algorithms perform well on a certain database, they do not achieve good results on other databases. To make a fair comparison, FERET evaluation (Phillips, Moon, Rizvi, & Rauss, 2000) and the Face Authentication Test (Messer et al., 2004) have been designed to evaluate different face identification and verification algorithms. However, these tests are not concerned with the speed of the algorithms. Since only accuracy is accounted for, the applicability of the algorithms to real-time applications is not considered. However, the trade-off between accuracy and speed is very important. In summary, a face recognition system should not only be able to cope with variations in illumination, expression and pose, but also recognize a face in real-time.

With in-plane face rotation, normalisation can be carried out using prominent facial features as a reference, e.g., the eyes. However, out-of plane rotation seems only to be solvable using 3D technologies. While the transformation of 3D data between different poses is trivial, 2D frontal view images can also be synthesized using a 3D model. The literature survey of 3D face model techniques in chapter 2 shows, however, that the process of synthesizing a frontal view image from an arbitrary pose using a 3D model is very slow. A number of approaches have also been proposed to use 3D data directly for recognition when such data is available. However, 3D scanners are still relatively expensive and there are still some significant limitations to be solved, e.g. the capture process is illumination sensitive, 3D depth resolution needs to be improved, etc. As a result, a 2D frontal view face recognition system is the main focus of this research.

Though quite a tough task for a computer, face recognition seems to be much easier for human beings. The ability to recognize faces and understand the emotions they convey is one of the most important human abilities. It is very common that one can instantly recognize thousands of people. Even a baby is able to identify its mother's face within

half an hour of birth. As with many perceptual abilities, the ease with which humans can recognize faces disguises the complexity of the task even when considering the many potential variations in such a dynamic real world object. An important outcome of research on artificial vision systems has shown that more than half of the cortex becomes more active during visual processing (Hallinan, Gordon, Yuille, Gibilin, & Mumford, 1999). The visual cortex thus plays a very important role in face recognition. Simple cells in the visual cortex are known to be selective for four coordinates, each cell having an *x, y* location in visual space, a preferred orientation and a preferred spatial frequency (Daugman, 1985). Based on this observation, a number of researches have actually shown that the various 2D receptive-field profiles encountered in populations of simple cells are well described by a family of 2D Gabor wavelets, which were first proposed by Gabor (1946) for simultaneous time and frequency analysis. In addition to this biological motivation, it is also widely believed that local texture features in face images, extracted by a spatial-frequency wavelet analysis, are basically more robust against distortions caused by various illumination, expression and pose (Zhao, Chellapa, Rosenfield, & Phillips, 2000). In particular, among various wavelet bases with good characteristics of space-frequency localization, the Gabor function provides the optimal resolution in both spatial and frequency domain (Gabor, 1946; Daugman, 1985). As a result, this research will apply 2D Gabor wavelets to extract features for face recognition. Since the simple cells of human visual cortex are well modelled and the local features in space and frequency domain are simultaneously extracted with optimal resolution, the system thus developed might be able to mimic a human's recognition ability and be more robust against the variation of illumination, expression and limited out of plane face rotation.

The motivation of this research is to develop both an accurate and a fast frontal view face recognition algorithm, which should be robust against variations in illumination, expression and limited out of plane face rotation. At the same time, the system will be efficient and applicable to real-time applications. When the recognition algorithm has been comprehensively tested against a number of different databases and its performance maximised, it will be implemented as a component of a fully automatic face recognition system, complete with face detection module.

1.6 Major Contributions of The Thesis

The major contributions of the thesis can be summarized as below:

- An overview on the background and applications of Gabor wavelet has been presented, which shows that this biologically driven mathematical tool can achieve the optimal resolution when performing joint time frequency analysis on the signal. The survey of applications of such wavelet to face recognition also provides some guidance for researchers in this area.
- A face recognition algorithm robust against variations of illumination, expression and limited out of plane rotations has been developed. Once Gabor features are extracted using a set of Gabor wavelets, kernel subspace methods are then applied to enhance classification accuracy. The algorithm is successfully applied to identification tasks and tested using public databases and protocols. The results verified the robustness of the extracted features against the nonlinear distortions caused by facial variations.
- Based on the successful application of Gabor features and kernel subspace methods to face identification, the method combining Generalized Discriminant Analysis (GDA) and Gabor features has also been successfully applied to verification. The

experimental results show that the algorithm is among the top performers in the Face Verification Competition 2004.

- A novel feature selection scheme, MutualBoost, has been proposed to learn the most important Gabor features for face recognition. The requirement of Gabor feature based methods for computation and memory can be substantially reduced when the selected features are used. The results show that MutualBoost selected Gabor features are more discriminative than those learned by the AdaBoost algorithm. The selected informative Gabor features are further combined with GDA (MutualGabor + GDA) for recognition and the method has been fully tested using the FERET database according to the evaluation protocol. The results show that MutualGabor + GDA achieves better performance than the top performer in the FERET evaluation, but with much higher efficiency.
- A novel symmetry based eye location method is presented in this research. By integrating the robust Gabor + GDA algorithm with the eye location method, a fully automatic verification system has been developed. When competing with 12 participants from around the world, the system ranked the 3rd in the Face Verification Competition (FVC2004).
- We have developed an automatic user identification system, which can effectively detect faces from a real time video stream, identify them and retrieve their registered personal information such as name etc. The system is expandable and fully integratable with other face detection and recognition algorithms.

1.7 Organization of the Thesis

The remaining chapters of this thesis are organized in the following way:

Chapter 2 introduces in detail the mathematical technologies used in the thesis. While Gabor wavelets are used for robust feature extraction, subspace analysis and support

vector machines are used for feature enhancement and classification. The AdaBoost algorithm and information theory are also described.

Chapter 3 reviews state of the art face recognition algorithms, both 2D based and 3D based approaches are included. Particularly, the major concern of the thesis, i.e., Gabor wavelet based methods are explored in detail.

Chapter 4 and 5 presents the proposed Gabor + GDA method for identification and verification, respectively. Both methodology and experimental results are given.

Chapter 6 describes a novel feature selection scheme and its application to select Gabor features for face recognition. The results show that the system using the selected Gabor features can significantly increase efficiency without deteriorating performance. In contrast, the face recognition system using the selected Gabor features has shown to be more robust against changes in illumination, pose and expression.

Chapter 7 proposes a generalized symmetry transform based eye location algorithm, which is tested using thousands of face images. The eye location module has also been integrated into an automatic verification algorithm and top performance on accuracy is observed when compared with other algorithms.

Chapter 8 presents an automatic user identification system developed in the research. Both system designs and function modules are explained.

Finally, chapter 9 gives conclusions and some comments for future research work on face recognition.

Chapter 2 Mathematical Techniques Used in This Thesis

This chapter is mainly concerned with the main mathematical techniques used in this thesis, which are listed as below:

- Gabor wavelets
- Linear Subspace Analysis
- Non-linear Kernel Subspace Analysis
- AdaBoost Learning Algorithm
- Support Vector Machine
- Entropy and Mutual Information

2.1 Joint Time Frequency Analysis and Gabor Wavelets

2.1.1 Joint Time Frequency Analysis and Gabor Function

For the past few decades the Fourier transform has been the most commonly used tool for signal frequency analysis (Ronald, 1978). It is, however, hard to tell where within a signal, certain frequencies occur, i.e., the information about the time domain is lost. Given the fact that the frequency content of the majority of signals in the real world change with time, it is far more useful to be able to characterize a signal in both the time and frequency domains simultaneously.

Instead of comparing the signal to complex sinusoidal functions, a natural way of representing a signal in time and frequency simultaneously is to compare the signal with elementary functions that are concentrated in both the time and frequency domains (Qian & Chen, 1996). Let $s(\tau)$ and $\varphi(\tau)$ be the signal and elementary function with centre frequency f, the joint time and frequency representation of the signal can thus be written as $\int s(\tau)\varphi(\tau - t)d\tau$, which is an inner product between the signal $s(\tau)$ and the shifted elementary function $\varphi(\tau)$. By moving the short time duration window function $\varphi(\tau)$, one could obtain information on how the signals frequency contents evolve over time. Suppose that the time duration and frequency bandwidth of $\varphi(\tau)$ are Δt and Δf respectively, then $\int s(\tau)\varphi(\tau - t)d\tau$ denotes signal information in the range of $[t - \Delta t, t + \Delta t] \times [f - \Delta f, f + \Delta f]$.

To achieve an exact measure of a signal at a particular time and frequency, Δt and Δf should be as narrow as possible. Unfortunately, the values of Δt and Δf are dependent on each other; they are related via the Fourier transform. It is well known that when the time duration increases, the frequency bandwidth must be smaller and vice versa (Ronald, 1978) thus there is always inherent uncertainty in the time and frequency

resolution of $\varphi(t)$. Several different methods are available to calculate the time duration and frequency bandwidth of a signal. The most common are the standard deviation, or root mean square (r.m.s.), this is a concept used in statistical theory (Qian et al., 1996; Daugman, 1985). The time duration Δt is defined as:

$$(\Delta t)^2 = \frac{\int_{-\infty}^{\infty} (t-\mu_t)^2 \varphi(t)\varphi^*(t)dt}{\int_{-\infty}^{\infty} \varphi(t)\varphi^*(t)dt}, \quad \mu_t = \frac{\int_{-\infty}^{\infty} t\varphi(t)\varphi^*(t)dt}{\int_{-\infty}^{\infty} \varphi(t)\varphi^*(t)dt} \tag{2.1.1}$$

By calculating the frequency uncertainty of Δf using a similar definition, it has been shown that there is a connection between the two uncertainties:

$$\Delta t\, \Delta f \geq \frac{1}{2} \tag{2.1.2}$$

Gabor (1946) derived the function that minimizes this uncertainty, i.e., turns the inequality into equalites such that $\Delta t\, \Delta f = \frac{1}{2}$. He found that the function is a Gaussian modulated by a sinusoidal signal:

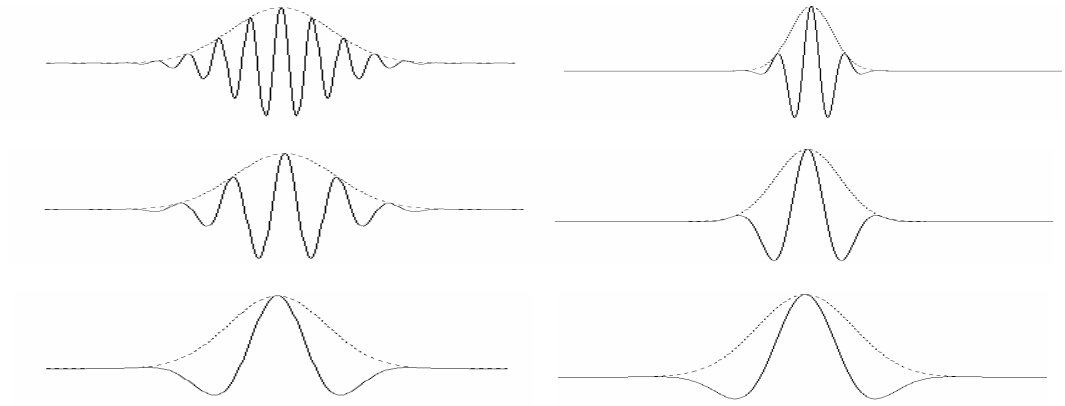
$$\varphi(t) = \exp(-\alpha^2 t^2)\exp(-j2\pi f_0 t) \tag{2.1.3}$$

where α is the sharpness of the Gaussian, and f_0 is the centre frequency of the sinusoidal signal. See Figure 2-1 for the Gabor elementary function with different frequencies. The function has a Fourier transform:

$$\Phi(f) = \sqrt{\frac{\pi}{\alpha^2}} \exp\left(-\frac{\pi^2}{\alpha^2}(f-f_0)^2\right) \tag{2.1.4}$$

As shown in Figure 2-1 (a), the shape of Gabor functions is decided by the Gaussian sharpness, which is invariant to the variance of the frequency. To make the time duration of function $\varphi(t)$ dependent on the central frequency f_0 (Daubechies, 1990; Kyrki, Kamarainen, & Kalviainen, 2004), a constant ratio $\gamma = \frac{f_0}{\alpha}$ is defined such that the function, when applied to different frequencies, behaves as a scaled version of each

other. Figure 2-1 (b) shows the Gabor functions with varied shape ($\gamma = \sqrt{2}$). Both the time duration and frequency bandwidth of the Gabor function are now related with the central frequency: the higher the frequency becomes, the smaller the time duration. This makes sense since high frequency signals change faster. The variations of time duration and frequency bandwidth in both domains are shown in Figure 2-2, which demonstrates the similarities between Gabor functions and other wavelets.



(a) (b)

Figure 2-1 Gabor elementary functions with fixed shape (a); with varied shapes (b)

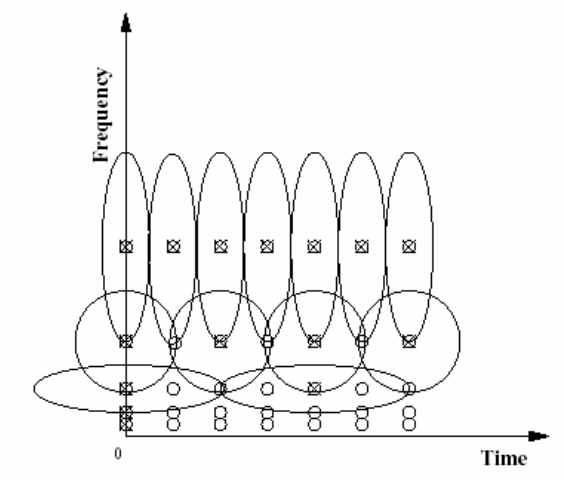


Figure 2-2 Time duration and frequency bandwidth of Gabor functions (Kyrki et al., 2004)

The maximum response of the function in the frequency domain can also be normalized to one by multiplying its inverse $\sqrt{\frac{\alpha^2}{\pi}}$. Consequently the normalized Gabor function is now defined as:

$$\varphi(t) = \frac{|f_0|}{\gamma\sqrt{\pi}} \exp(-\frac{f_0^{\ 2}}{\gamma}\ t^2) \exp(-j2\pi f_0 t) \qquad (2.1.5)$$

2.1.2 2D Gabor Wavelets

The 2D counterpart of a Gabor elementary function was first introduced by Granlund (1978). It can be derived directly from (2.1.5) by replacing t with spatial coordinates (x, y). Daugman (1985) showed a surprising equivalence between the 2D Gabor function and the organization and characteristics of the mammalian visual system. By generalizing the time frequency resolution uncertainty to the 2D domain, i.e., $\Delta x \Delta y \Delta u \Delta v \geq \frac{1}{4}$, he also showed that the joint 2D resolution of Gabor wavelets actually achieves the theoretical limit regardless of the values of any of the parameters. From an information theoretic viewpoint, Okajima (1998) derived the Gabor functions as solutions for a certain mutual-information maximization problem. The work shows that the Gabor-type receptive field can extract the maximum information from local image regions. Setting the sharpness of the Gaussian in the y axis as β and the ratio with the central frequency as $\eta = \frac{f}{\beta}$, the 2D Gabor wavelet can now be defined as (Kyrki et al., 2004):

$$\varphi(x, y) = \frac{f^2}{\pi\gamma\eta} \exp\left(-(\alpha^2 x_r^{\ 2} + \beta^2 y_r^{\ 2})\right) \exp\left(j2\pi f x_r\right) \qquad (2.1.6)$$
$$x_r = x\cos\theta + y\sin\theta, y_r = -x\sin\theta + y\cos\theta$$

where f is the frequency of the modulating sinusoidal plane wave and θ is the orientation of the major axis of the elliptical Gaussian. The 2D Gabor wavelet as defined in (2.1.6) has the Fourier transform:

$$\Phi(u,v) = \exp\left(-\pi^2\left(\frac{1}{\alpha^2}(u_r - f)^2 + \frac{1}{\beta^2}v_r^2\right)\right) \qquad (2.1.7)$$
$$u_r = u\cos\theta + v\sin\theta, v_r = -u\sin\theta + v\cos\theta$$

The plots for two Gabor wavelets in the spatial and frequency domains are shown in Figure 2-3.

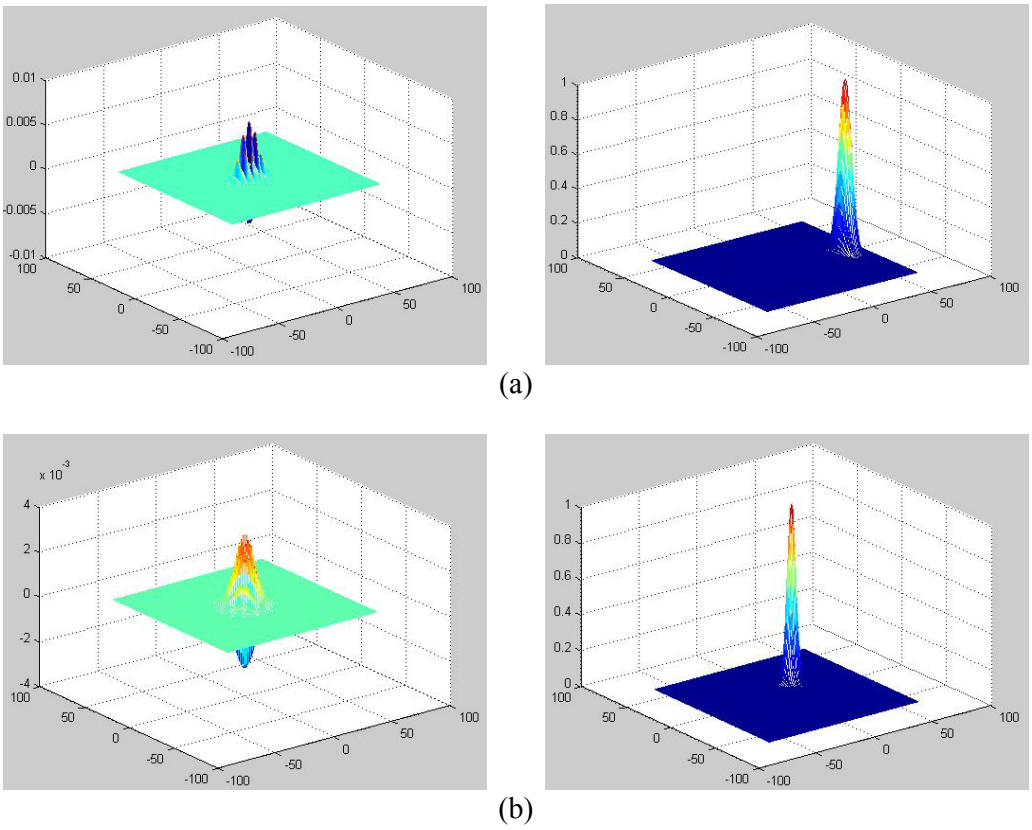


Figure 2-3 Example 2D Gabor wavelets in the spatial and the frequency domain (a) $f = 0.4, \theta = 0, \gamma = 4, \eta = 2$ **, (b)** $f = 0.2, \theta = \pi/4, \gamma = 2, \eta = 2$

Note that the equation defined in (2.1.6) is different to the one normally used for face recognition (Lades et al., 1993; Wiskott, Fellous, Kruger, & von der Malsburg, 1997; Liu & Wechsler, 2002), however, this definition is more general. To find the relationship between different Gabor wavelet definitions, we firstly define a wave vector $\vec{k} = 2\pi f \exp(j\theta)$ to represent the central frequency components in the frequency domain. Note the assumption here is that the orientation of the wave vector is the same

as that of major axis of the elliptical Gaussian, which is fully supported by the models of receptive fields found in simple cells of the cat and macaque striate cortices (Daugman, 1985; Jones & Palmer, 1987). Setting $\gamma = \eta = \frac{\sigma}{\sqrt{2}\pi}$, i.e. $\alpha = \beta = \frac{\sqrt{2}\pi f}{\sigma}$, the Gabor wavelet located at position $\vec{z} = (x, y)$ can now be defined as:

$$\varphi(\vec{z}) = \frac{1}{2\pi}\frac{\left\|\vec{k}\right\|^2}{\sigma^2}\exp\left(\frac{-\left\|\vec{k}\right\|^2\left\|\vec{z}\right\|^2}{2\sigma^2}\right)\exp(i\vec{k}\cdot\vec{z}) \qquad (2.1.8)$$

The wavelet function used in (Lades et al., 1993; Wiskott et al., 1997; Liu et al., 2002) has thus been derived from equation (2.1.6), and can been seen as a special case with $\alpha = \beta$. Similarly, the relationship between equation (2.1.6) and those defined in (Fasel, Barlett, & Movellan, 2002; Weldon, Higgins, & Dunn, 1996) could also be established, where the DC term could be deduced to make the wavelet DC free (Lades et al., 1993; Wiskott et al., 1997; Liu et al., 2002), similar effects can also be achieved by normalizing the image to be zero mean (Kruger & Sommer, 2000; Kruger & Sommer, 2002a).

2.2 Linear Subspace Analysis

2.2.1 Principal Component Analysis (PCA)

The aim of PCA is to identify a subspace spanned by the training images $\{x_1, x_2, \cdots x_i \cdots x_M\}$, which could decorrelate the variance of pixel values. This can be achieved by eigen analysis of the covariance matrix $\mathbf{\Sigma} = \frac{1}{M-1}\sum_{i=1}^{M}(x_i - \bar{x})(x_i - \bar{x})^T$:

$$\mathbf{\Sigma E} = \mathbf{\Lambda E} \qquad (2.2.1)$$

where $\mathbf{E}$, $\mathbf{\Lambda}$ are the resultant eigenvectors, also referred to as eigenfaces, and eigen values respectively. The representation of a face image in the PCA subspace is then

obtained by projecting it to the coordinate system defined by the eigenfaces (Turk & Pentland, 1991).

2.2.2 Linear Discriminant Analysis (LDA)

While the projection of face images into PCA subspace achieves decorrelation and dimensionality reduction, LDA aims to find a projection matrix $\mathbf{W}$ which maximizes the quotient of the determinants of $\mathbf{S}_b$ and $\mathbf{S}_w$ (Zhao, Krishnaswamy, Chellapa, Swets, & Weng, 1998),

$$\mathbf{W} = \arg\max \frac{\left|\mathbf{W}^T \mathbf{S}_b \mathbf{W}\right|}{\left|\mathbf{W}^T \mathbf{S}_w \mathbf{W}\right|} \tag{2.2.2}$$

where $\mathbf{S}_b$ and $\mathbf{S}_w$ are the between-class scatter and within-class scatter respectively. Consider a C class problem and let N_c be the number of samples in class c, a set of M training patterns from the C class can be defined as $\{x_{ck}, c = 1,2,...C; k = 1,2,..., N_c\}, M = \sum_{c=1}^{C} N_c$. The S_b and S_w of a training set can be computed as:

$$\mathbf{S}_w = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{N_c}\sum_{k=1}^{N_c}(x_{ck} - \mu_c)(x_{ck} - \mu_c)^T \tag{2.2.3}$$

$$\mathbf{S}_b = \frac{1}{C}\sum_{c=1}^{C}(\mu_c - \mu)(\mu_c - \mu)^T \tag{2.2.4}$$

where μ is the mean of the whole training set, and μ_c is the mean for the class c. It was shown in (Fukunnaga, 1991) that the projection matrix $\mathbf{W}$ can be computed from the eigenvectors of $\mathbf{S}_w^{-1}\mathbf{S}_b$. However, due to the high dimensionality of the feature vector, especially in face recognition applications, $\mathbf{S}_w$ is usually singular, i.e. the inverse of $\mathbf{S}_w$ does not exist. As a result, a two-stage dimensionality reduction technique, named the Most Discrimiant Features (MFD), was proposed by (Swets & Weng, 1996). The original face vectors are first projected to a lower dimensional space by PCA, which is then subjected to LDA analysis. Let $\mathbf{W}_{pca}$ be the projection matrix from the original

image space to the PCA subspace, the LDA projection matrix W_{lda} is thus composed of the eigenvectors of $(\mathbf{W}_{pca}{}^{T}\mathbf{S}_{w}\mathbf{W}_{pca})^{-1}(\mathbf{W}_{pca}{}^{T}\mathbf{S}_{b}\mathbf{W}_{pca})$. The final projection matrix $\mathbf{W}_{mfd}$ can thus be obtained by:

$$\mathbf{W}_{mfd} = \mathbf{W}_{pca} \times \mathbf{W}_{lda} \tag{2.2.5}$$

Note that the rank of $\mathbf{S}_{b} \le C-1$, while the rank of $\mathbf{S}_{w} \le M-C$. As a result, it is suggested that the dimension of the PCA subspace should be *M*-*C* (Swets et al., 1996).

2.3 Non-linear Kernel Subspace Analysis

As seen from last section, both PCA and LDA are linear methods. Since facial variations are mostly nonlinear, PCA and LDA projections could only provide suboptimal solutions for face recognition tasks (Gupta & Agrawal, 2002). Recently, kernel methods have been successfully applied to solve pattern recognition problems because of their capacity in handling nonlinear data. Support Vector Machines (SVMs) are typical kernel methods and have been successfully applied to face detection (Osuna, Freund, & Girosit, 1997), face recognition (Phillips, 1999) and gender classification (Moghaddam & Yang, 2000). By mapping sample data to a higher dimensional feature space, effectively a nonlinear problem defined in the original image space is turned into a linear problem in the feature space (Scholkopf et al., 1999). PCA or LDA can subsequently be performed in the feature space and thus Kernel Principal Component Analysis (KPCA) (Scholkopf, Smola, & Muller, 1998) and Generalized Discrimniant Analysis (GDA) (Baudat & Anouar, 2000). Experiments show that KPCA and GDA are able to extract nonlinear features and thus provide better recognition rates in applications such as character (Scholkopf et al., 1998) and face recognition (Kim, Jung, & Kim, 2002; Yang, 2002).

2.3.1 The Kernel Feature Space

Algorithms in feature spaces make use of the following idea: via a nonlinear mapping

$$\begin{aligned} &\phi : R^N \to F \\ &x \to \phi(x) \end{aligned} \tag{2.3.1}$$

the data $\{x_k \in R^N, k = 1,...,M\}$ is mapped into a potentially much higher dimensional feature space F. Classification may be much easier in this feature space since a simple linear classifier will be adequate. Intuitively, the idea can be understood from the simple example in Figure 2-4. While a complicated nonlinear decision surface is needed in the two dimensional space, a simple hyper-plane is enough in the mapped feature space to separate the classes:

$$\begin{aligned} &\phi : R^2 \to R^3 \\ &(x_1, x_2) \to (z_1, z_2, z_3) = (x_1^2, \sqrt{2}x_1x_2, x_2^2) \end{aligned} \tag{2.3.2}$$

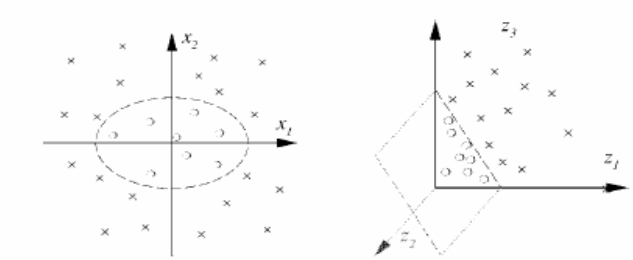


Figure 2-4 A simple example (2D->3D) (Muller, Mika, Ratsch, Tsuda, & Scholkopf, 2001)

In this example, the complexity of algorithms can be easily controlled due to the low dimension of feature space. However, when the dimension of feature space is huge, e.g. image related classification problems, it would be intractable to execute an algorithm in this space. Fortunately, there is a highly effective trick for computing dot products in feature spaces for certain mappings ϕ and feature spaces F: kernel functions (Scholkopf et al., 1999) . In the simple example, the dot product between two feature space vectors can be easily computed with a kernel function *k* as below:

$$\begin{aligned}(\phi(x)\cdot\phi(y)) &= (x_1^2, \sqrt{2}x_1x_2, x_2^2)\cdot(y_1^2, \sqrt{2}y_1y_2, y_2^2)\\ &= ((x_1, x_2)\cdot(y_1, y_2))^2\\ &= (x\cdot y)^2\\ &= k(x, y)\end{aligned} \tag{2.3.3}$$

There exists a feature space F and mapping ϕ such that $k(x,y) = (\phi(x)\cdot\phi(y))$, if the function $k(x,y)$ satisfies Mercer's condition (Scholkopf et al., 1999). The most widely used kernel functions are the Polynomial kernel $k(x,y) = (x\cdot y)^d$ and the RBF kernel $k(x,y) = \exp\left(-\|x-y\|^2 \Big/ r\right)$.

2.3.2 Kernel Principal Component Analysis (KPCA)

Suppose the training patterns in the input space R^N are $\{x_k, k = 1,...,M\}$. ϕ is the non-linear map defined from the input space to a high dimensional feature space: $\phi: R^N \to F$. Each vector x_k is now mapped to a higher dimension vector $\phi(x_k)$ in the feature space. Here, we assume all the data mapped into the feature space are centred, i.e.

$$\sum_{k=1}^{M}\phi(x_k) = 0 \tag{2.3.4}$$

The covariance matrix of the training samples in the feature space is now:

$$\mathbf{C} = \frac{1}{M}\sum_{k=1}^{M}\phi(x_k)\phi(x_k)^T \tag{2.3.5}$$

Kernel PCA aims to find the eigenvalues $\lambda \geq 0$ and eigenvectors $v \in F \setminus \{0\}$ satisfying

$$\lambda\mathbf{v} = \mathbf{C}\mathbf{v} \tag{2.3.6}$$

All solutions **v** lie in the span of $\phi(x_1)$, …, $\phi(x_m)$, and there exist coefficients α_k such that

$$\mathbf{v} = \sum_{k=1}^{M}\alpha_k\phi(x_k) \tag{2.3.7}$$

Take the inner-product with vector $\phi(x_k)(k = 1,...,M)$ on both sides of (2.3.6):

$$\lambda(\mathbf{v}\cdot\phi(x_k)) = (\mathbf{C}\mathbf{v})\cdot\phi(x_k) \tag{2.3.8}$$

By substituting (2.3.5) and (2.3.7) into (2.3.8) and defining a $M \times M$ matrix $\mathbf{K}$ with:

$$\mathbf{K}_{ij} = k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \tag{2.3.9}$$

the following can be obtained:

$$M\lambda \mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha} \Rightarrow M\lambda\boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha} \tag{2.3.10}$$

where $\boldsymbol{\alpha}$ denotes a column vector with entries $\alpha_1,...,\alpha_M$. The above derivation assumes that all the mapped data $\phi(x_k)$ is centred in feature space F. See section 2.3.4 for an approach to centre the data $\phi(x_k)$ in F.

For a new pattern x, the projection of its image $\phi(x)$ in the feature space onto the eigenvector $\mathbf{v}$ can now be computed as:

$$\mathbf{v} \cdot \phi(x) = \sum_{k=1}^{M} \alpha_k (\phi(x_k) \cdot \phi(x)) = \sum_{k=1}^{M} \alpha_k k(x_k, x) \tag{2.3.11}$$

If the first L $(1 \le L \le M)$ significant eigenvectors are extracted to construct the eigen matrix:

$$\mathbf{W} = [\boldsymbol{\alpha}_1\ \boldsymbol{\alpha}_2 \ldots \boldsymbol{\alpha}_L] \tag{2.3.12}$$

The projection of x in the L-dimensional Kernel PCA space is given by:

$$\mathbf{y} = \mathbf{k}_x \mathbf{W} \tag{2.3.13}$$

where

$$\mathbf{k}_x = [k(x, x_1)\, k(x, x_2) \ldots k(x, x_M)] \tag{2.3.14}$$

2.3.3 Generalized Discriminant Analysis

As a generalized version of Linear Discriminant Analysis (LDA), Generalized Discriminant Analysis (GDA) performs LDA on sample data in the high dimension feature space. Consider a C class problem and let N_c be the number of samples in class c, then the set of training patterns in class c can be defined as $\{x_{ck}, c = 1,2,...C; k = 1,2,..., N_c\}$. The total number of input vectors can be denoted as:

$M = \sum_{c=1}^{C} N_c$. For a centred data set in feature space, the between-class scatter matrix $\mathbf{S}_b$ and within-class scatter matrix $\mathbf{S}_w$ can be defined as:

$$\mathbf{S}_w = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{N_c}\sum_{k=1}^{N_c}\phi(x_{ck})\phi(x_{ck})^T \tag{2.3.15}$$

$$\mathbf{S}_b = \frac{1}{C}\sum_{c=1}^{C}\mu_c\mu_c^T \tag{2.3.16}$$

where μ_c is the mean vector of class c:

$$\mu_c = \frac{1}{N_c}\sum_{k=1}^{N_c}\phi(x_{ck}) \tag{2.3.17}$$

Similar to LDA, the purpose of GDA is to maximize the quotient between the inter-class inertia and the intra-classes inertia. This maximization is equivalent to finding eigenvalues $\lambda \geq 0$ and eigenvectors $\mathbf{v} \in F \setminus \{0\}$ satisfying

$$\lambda\mathbf{S}_w\mathbf{v} = \mathbf{S}_b\mathbf{v}\,, \tag{2.3.18}$$

all solutions $\mathbf{v}$ lie in the span of $\phi(x_{11})$, …, $\phi(x_{ck})$, … and there exist coefficients α_{ck} such that

$$\mathbf{v} = \sum_{c=1}^{C}\sum_{k=1}^{N_c}\alpha_{ck}\phi(x_{ck}) \tag{2.3.19}$$

Substitute (2.3.15), (2.3.16), (2.3.17) and (2.3.19) into (2.3.18):

$$\frac{\lambda}{C}\frac{1}{N_p}\sum_{q=1}^{C}\sum_{j=1}^{N_q}\alpha_{qj}\sum_{p=1}^{C}\sum_{i=1}^{N_p}\phi(x_{pi})\phi(x_{pi})^T\phi(x_{qj}) = \frac{1}{C}\sum_{q=1}^{C}\sum_{j=1}^{N_q}\alpha_{qj}\sum_{p=1}^{C}\left[\frac{1}{N_p}\sum_{i=1}^{N_p}\phi(x_{pi})\right]\left[\frac{1}{N_p}\sum_{i=1}^{N_p}\phi(x_{pi})^T\phi(x_{qj})\right] \tag{2.3.20}$$

Now take the inner-product with vector $\phi(x_{ck})$ on both sides:

$$\lambda\sum_{q=1}^{C}\sum_{j=1}^{N_q}\alpha_{qj}\sum_{p=1}^{C}\sum_{i=1}^{N_p}\left(\phi(x_{pi})\cdot\phi(x_{ck})\right)\left(\phi(x_{pi})\cdot\phi(x_{qj})\right) = \sum_{q=1}^{C}\sum_{j=1}^{N_q}\alpha_{qj}\sum_{p=1}^{C}\left[\sum_{i=1}^{N_p}\left(\phi(x_{pi})\cdot\phi(x_{ck})\right)\right]\left[\frac{1}{N_p}\sum_{i=1}^{N_p}\left(\phi(x_{pi})\cdot\phi(x_{qj})\right)\right] \tag{2.3.21}$$

The dot product of a sample *i* from class *p* and the other sample *j* from class *q* in the feature space, denoted as $(k_{ij})_{pq}$, can be calculated by a kernel function, e.g., radial basis kernels as below:

$$(k_{ij})_{pq} = \phi(x_{pi}) \cdot \phi(x_{qj}) = k(x_{pi}, x_{qj}) = e^{-\|x_{pi} - x_{qj}\|^2 / r} \tag{2.3.22}$$

Let $\mathbf{K}$ be a $M \times M$ matrix defined on the class elements by $((\mathbf{K}_{pq})_{\substack{p=1,\ldots C \\ q=1,\ldots C}})$, where $\mathbf{K}_{pq}$ is a matrix composed of dot products between vectors from class *p* and *q* in feature space:

$$\mathbf{K}_{pq} = (k_{ij})_{\substack{i=1,\ldots,N_p \\ j=1,\ldots,N_q}} \tag{2.3.23}$$

Also define a $M \times M$ block diagonal matrix:

$$\mathbf{U} = (\mathbf{U}_c)_{c=1,\ldots,C} \tag{2.3.24}$$

where $\mathbf{U}_c$ is $N_c \times N_c$ a matrix with terms all equal to $\frac{1}{N_c}$.

The equation in (2.3.21) can now be represented as:

$$\lambda \mathbf{K}\mathbf{K}\boldsymbol{\alpha} = \mathbf{K}\mathbf{U}\mathbf{K}\boldsymbol{\alpha} \tag{2.3.25}$$

where $\boldsymbol{\alpha}$ denotes a column vector with M entries $\alpha_{ck}, c = 1,\ldots C, k = 1,\ldots, N_c$. Different techniques can be used to solve the eigen problem given in (2.3.25), the algorithm proposed by (Baudat et al., 2000) was adopted in this thesis, which finds the eigen vector $\mathbf{v}$ by first diagonalizing the matrix $\mathbf{K}$. Once $\boldsymbol{\alpha}$ and the eigenvectors $\mathbf{v}$ are decided upoon, the projection of a new sample x in the GDA space can be easily calculated using equations (2.3.12)-(2.3.14). Details on eigenvalue resolutions of GDA can be found in Appendix A.

2.3.4 Non-centred Data

In the general case, data $\{\phi(x_i)\}, i = 1,2,\cdots,M$ is not centred in the feature space. The following method can be used make this datas mean zero:

$$\tilde{\phi}(x_i) = \phi(x_i) - \frac{1}{M}\sum_{m=1}^{M}\phi(x_m) \tag{2.3.26}$$

The $M \times M$ kernel matrix $\tilde{\mathbf{K}}$ for the centred data can now be calculated as:

$$\begin{aligned}
\left(\tilde{\mathbf{K}}\right)_{ij} &= \tilde{\phi}(x_i)\cdot\tilde{\phi}(x_j) \\
&= \left(\phi(x_i) - \frac{1}{M}\sum_{m=1}^{M}\phi(x_m)\right)\cdot\left(\phi(x_j) - \frac{1}{M}\sum_{\mathrm{n}=1}^{M}\phi(x_{\mathrm{n}})\right) \\
&= \left(\phi(x_i)\cdot\phi(x_{\mathrm{J}})\right) - \frac{1}{M}\sum_{m=1}^{M}\phi(x_m)\cdot\phi(x_j) - \frac{1}{M}\sum_{\mathrm{n}=1}^{M}\phi(x_i)\cdot\phi(x_{\mathrm{n}}) + \frac{1}{M^2}\sum_{m=1}^{M}\sum_{\mathrm{n}=1}^{M}\phi(x_m)\cdot\phi(x_{\mathrm{n}}) \\
&= \left(\mathbf{K}\right)_{ij} - \frac{1}{M}\sum_{m=1}^{M}\left(\mathbf{K}\right)_{\mathrm{mj}} - \frac{1}{M}\sum_{\mathrm{n}=1}^{M}\left(\mathbf{K}\right)_{\mathrm{in}} + \frac{1}{M^2}\sum_{m=1}^{M}\sum_{\mathrm{n}=1}^{M}\left(\mathbf{K}\right)_{\mathrm{mn}}
\end{aligned} \tag{2.3.27}$$

Which, represented in matrix form is as follows:

$$\tilde{\mathbf{K}} = \mathbf{K} - \frac{1}{M}\mathbf{1}_M\mathbf{K} - \frac{1}{M}\mathbf{K}\mathbf{1}_M + \frac{1}{M^2}\mathbf{1}_M\mathbf{K}\mathbf{1}_M \tag{2.3.28}$$

Once the kernel matrix $\tilde{\mathbf{K}}$ for the centred data is calculated, the same procedures as used in previous sections can be used to compute the projection matrix $\mathbf{W}$ for the KPCA, or GDA subspace. As given in Equation 2.3.13, the projection of a new pattern x into the learned subspace can now be computed as:

$$\mathbf{y} = \tilde{\mathbf{k}}_{\mathbf{x}}\mathbf{W} \tag{2.3.29}$$

where

$$\begin{aligned}
\left(\tilde{\mathbf{k}}_{\mathbf{x}}\right)_i &= \tilde{\phi}(x)\cdot\tilde{\phi}(x_i) \\
&= \left(\phi(x) - \frac{1}{M}\sum_{m=1}^{M}\phi(x_m)\right)\cdot\left(\phi(x_i) - \frac{1}{M}\sum_{\mathrm{n}=1}^{M}\phi(x_{\mathrm{n}})\right) \\
&= k(x,x_i) - \frac{1}{M}\sum_{m=1}^{M}k(x_i,x_m) - \frac{1}{M}\sum_{n=1}^{M}k(x,x_n) + \frac{1}{M^2}\sum_{m=1}^{M}\sum_{\mathrm{n}=1}^{M}k(x_m,x_{\mathrm{n}})
\end{aligned} \tag{2.3.30}$$

Next we define a $1\times M$ row vector $\mathbf{1}$ with all entries equal to 1, the equation can then be represented in matrix form:

$$\tilde{\mathbf{k}}_{\mathbf{x}} = \mathbf{k}_{\mathbf{x}} - \frac{1}{M}\mathbf{1}\mathbf{K} - \frac{1}{M}\mathbf{k}_{\mathbf{x}}\mathbf{1}_M + \frac{1}{M^2}\mathbf{1}\mathbf{K}\mathbf{1}_M \tag{2.3.31}$$

2.4 AdaBoost Learning Algorithm

Introduced by Freud and Schapire (1999), AdaBoost has been successfully applied to object detection (Viola & Jones, 2001; Lienhart & Maydt, 2002) and face recognition

(Michael & Viola, 2003). The essence of AdaBoost is to learn a number of very simple weak classifiers, which are then linearly combined into a single strong classifier. Whilst the performance of weak classifiers could be just slightly better than random guessing, AdaBoost learning minimizes the upper bound on both training and generalization errors (Freund et al., 1999). Additionally AdaBoost has been applied to select Haar-like features (Lienhart et al., 2002) for face detection, recognition (Michael et al., 2003) and Gabor feature selection (Shen & Bai, 2004a) for classification.

Given *m* training samples $(x_i, y_i), i = 1,2,..,m, x_i \in R^N, y_i \in \{-1,1\}$
Initialization: weights $w_1(i) = 1/m$
For *t=1, ..., T*

1) Train weak learners using distribution w_t
2) Choose a weak hypothesis $h_t : R^n \to \{-1,1\}$
3) Choose $\alpha_t \in R$
4) Update weights: $w_{t+1}(i) = \frac{w_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}$

Final strong classifier: $H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$

Figure 2-5 Details of AdaBoost algorithm (Freund et al., 1999)

2.4.1 The Algorithm

For two class problems, a set of *m* labelled training samples is given as $(x_i, y_i), i = 1,2,..,m$, where $y_i \in \{-1,1\}$ is the class label associated with sample $x_i \in R^n$. A large number of weak classifiers $h : R^n \to \{-1,1\}$ could be generated to form the classifier pool for learning. The weak classifier could be very simple, e.g., a threshold function on the *kth* coordinate of x in the *n*-dimensional space. The algorithm focuses on the difficult training patterns, increasing their representation in successive training sets. Over a number of *T* rounds, *T* weak classifiers are selected to form the final strong classifier. In each of the iterations, the space of all possible weak classifiers is searched exhaustively to find the one with the lowest weighted classification error. The error is then used to update the weights such that the wrongly classified samples get their weights increased.

The resulting strong classifier is a weighted linear combination of all T selected weak classifiers. Figure 2-5 contains the listing of the AdaBoost algorithm.

2.4.2 Training Error

Letting $f(x)=\sum_{t=1}^{T}\alpha_t h_t(x)$ and unravelling the weight update rule:

$$\begin{aligned} w_{T+1}(i) &= \frac{w_T(i)\exp\left(-\alpha_t y_i h_t(x_i)\right)}{Z_T} \\ &= \frac{\exp\left(-\sum_t \alpha_t y_i h_t(x_i)\right)}{m\prod_t Z_t} \\ &= \frac{\exp\left(-y_i f(x_i)\right)}{m\prod_t Z_t} \end{aligned} \qquad (2.4.1)$$

Also let $[\pi]$ be an indicator variable which is 1 if the predicate π is true and 0 otherwise. Moreover, if $H(x_i)\neq y_i$ then $y_i f(x_i)\leq 0$ implying that $\exp(-y_i f(x_i))\geq 1$. Thus,

$$[H(x_i)\neq y_i]\leq \exp(-y_i f(x_i)) \qquad (2.4.2)$$

Since the training error of H : $\varepsilon_{tr}(H)$ is simply the number of wrongly classified samples divided by m, the bound of the error can be easily found as below:

$$\begin{aligned} \varepsilon_{tr}(H) &= \frac{1}{m}\sum_i [H(x_i)\neq y_i] \\ &\leq \frac{1}{m}\sum_i \exp(-y_i f(x_i)) \\ &= \frac{1}{m}\sum_i \left(m w_{T+1}(i)\prod_t Z_t\right) \\ &= \prod_t Z_t \end{aligned} \qquad (2.4.3)$$

2.4.3 Choosing α_t and h_t

To make w_{t+1} be a distribution, the value of Z_t shall actually be the sum of $w_{t+1}(i)$:

$$Z_t=\sum_i w_{t+1}(i)=\sum_i w_t(i)\exp(-\alpha_t y_i h_t(x_i)) \qquad (2.4.4)$$

To minimize the upper bound of training error: $\prod_t Z_t$, a greedy algorithm chooses α_t and h_t such that Z_t is minimized on each round of training. By using a linear upper

bound function $Z(\alpha_t)$ of Z_t and setting the derivative $dZ/d\alpha_t$ to zero, the value of α_t to minimize Z_t is found to be (see appendix B for details):

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1+r_t}{1-r_t}\right) \tag{2.4.5}$$

where $r_t = \sum_i w_t(i) y_i h_t(x_i)$. Since Z_t is now bounded by $Z_t \le \sqrt{1-r_t^2}$, the training error of H is now at most $\prod_t \sqrt{1-r_t^2}$. The training error can be further minimized if h_t is chosen such that $|r_t|$ is maximized on each round of boosting. Since r_t is closely related with the prediction error ε_t of h_t as below (see appendix B for details):

$$\varepsilon_t = \sum_i w_t(i)\left[h_t(x_i) \neq y_i\right] = \frac{1-r_t}{2}, \tag{2.4.6}$$

maximizing r_t is equivalent to minimizing error ε_t. In sum, h_t with minimum prediction error ε_t should be chosen on each round of boosting and α_t should be set as:

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) \tag{2.4.7}$$

2.5 Support Vector Machine

Originating from the hyperplane classifier proposed by (Boser, Guyon, & Vapnik, 1992), the support vector machine (SVM) has been greatly developed and widely applied in machine learning, classification and pattern recognition ever since (Scholkopf et al., 1997; Cristianini Nello & Shawe-Taylor John, 2000; Moghaddam et al., 2000; Osuna et al., 1997).

The SVM is basically a linear hyperplane classifier $f(x) = \langle w, x \rangle + b$ aimed at solving the two class problem. As shown in Figure 2-6, the classifier can separate the data from two classes very well. Since there might be a number of such linear classifiers available, SVM chooses the one with the maximal margin, which is defined as the width that the boundary could be increased by before hitting a data point. The distance between the

two thin lines (boundary) in the figure thus defines the margin of the linear SVM with data points on the boundary known as support vectors. The linear classifier $f(x)$ with maximized margin can be found using quadratic programming (QP) optimisation techniques as below:

$$f(x) = sign\left(\sum \alpha_k y_k \langle x_k, x \rangle + b\right) \qquad (2.5.1)$$

where $x_k \in R^N$ is the support vectors learned by SVM.

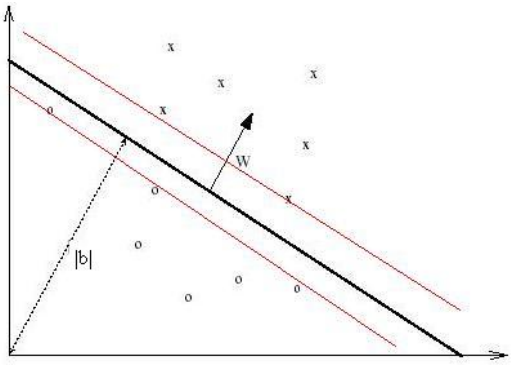


Figure 2-6 A hyperplane classifier in 2-dimension feature space

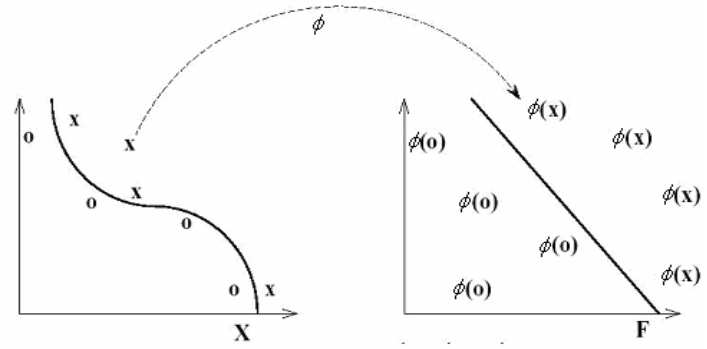


Figure 2-7 Map data into a feature space where they are linearly separable

When the data is non-separable, by relaxing constraints and introducing extra error to the objective function, linear SVM can also be solved using QP techniques. For non-linearly separable data, a nonlinear mapping function $\phi : R^N \to F, x \to \phi(x)$ is used to map them into a higher dimension feature space where the linear classifier can be applied. Figure 2-7 shows an example using the kernel method to learn non-linear SVM, which

is similar to the one shown in Figure 2-4. Using the same kernel trick as described in section 3, the non-linear SVM is now found to be:

$$f(x) = sign\left(\sum \alpha_k y_k k(x_k, x) + b\right) \qquad (2.5.2)$$

where $k(x_k, x)$ is a kernel function, e.g., a polynomial kernel and a Gaussin kernel etc. Given a set of training samples $(x_1, y_1),...,(x_i, y_i),...,(x_l, y_l), x_i \in R^N, y_i \in \{-1,+1\}$, SVM not only achieved only small error on the training set, it also minimized the upper bound of the error on a test set, i.e. generalization error (Burges, 1998). It has been shown by researchers that, with probability $1-\eta, 0 \le \eta \le 1$, the following bound on the expected generalization error of the SVM holds:

$$R < R_{emp} + \sqrt{\frac{h(\log(2l/h)+1) - \log(\eta/4)}{l}} \qquad (2.5.3)$$

where $R_{emp} = \frac{1}{l}\sum_{i=1}^{l} |f(x_i) - y_i|$ is the empirical risk as measured on the training set and h is the Vapnik Chervonenkis (VC) dimension. The second term on the right hand side is called the VC confidence. The SVM minimises the upper bound by fixing the empirical risk to a small value and minimising the VC confidence.

2.6 Entropy and Mutual Information

As a basic concept in information theory, entropy $H(X)$ is used to measure the uncertainty of a random variable (r.v.) X. If X is a discrete r.v., $H(X)$ can be defined as below:

$$H(X) = -\sum_x p(X = x) \lg(p(X = x)) \qquad (2.6.1)$$

Mutual information $I(Y;X)$ is a measure of general interdependence between two random variables X and Y:

$$I(Y;X) = H(X) + H(Y) - H(X,Y) \qquad (2.6.2)$$

Using Bayes rule on conditional probabilities, Eq. (2.6.2) can be rewritten as:

$$I(Y;X) = H(X) - H(X|Y) = H(Y) - H(Y|X) \qquad (2.6.3)$$

Since $H(Y)$ measures the priori uncertainty of Y and $H(Y|X)$ measures the conditional posterior uncertainty of Y after X is observed, the mutual information $I(Y;X)$ measures how much the uncertainty of Y is reduced if X has been observed. It can be easily shown that if X and Y are independent, $H(X,Y) = H(X) + H(Y)$, consequently their mutual information is zero.

2.7 Notation Definitions

2.7.1 Gabor Jet and Similarity Function

The convolution of an image *I* and a 2D Gabor wavelet φ can be defined as follows:

$$G(\vec{x}) = I * \varphi(\vec{x}), \vec{x} = (x, y) \qquad (2.7.1)$$

where $G(\vec{x})$ denote the convolution result with a wavelet at a position $\vec{x}$. Since the local frequency and orientation information is not available, a number of, e.g. 40, wavelets $\varphi_j, j = 0,1,...,39$ tuned to different frequencies and orientations are normally used for feature extraction. The convolution results at a position $\vec{x}$ thus consist of important local information, and can be concatenated to form a discriminative local feature, i.e. jet. A jet $J(\vec{x})$ is defined as the set of 40 complex coefficients $\{J_j(\vec{x}), j = 0,1,...,39\}$ obtained at one image point $\vec{x}$, where $J_j(\vec{x}) = I * \varphi_j(\vec{x})$. The complex coefficient J_j can also be written as $J_j = a_j \exp(i\phi_j)$ with magnitude a_j and phase ϕ_j, which contains very important local texture information.

Two functions $S_m(J, J')$ and $S_p(J, J')$ are defined to measure the similarity between two jets J and J'. While the first function S_m uses the magnitude information only, the other function S_p takes phase information into consideration as well. The two similarity functions are defined as below:

$$S_m(J,J') = \frac{J \cdot J'}{\|J\|\|J'\|} = \frac{\sum_j a_j a'_j}{\sqrt{\sum_j a_j \sum_j a'_j}} \quad (2.7.2)$$

$$S_p(J,J') = \frac{\sum_j a_j a'_j \cos(\phi_j - \phi'_j - \vec{d}\vec{k}_j)}{\sqrt{\sum_j a_j^2 \sum_j a'^2_j}} \quad (2.7.3)$$

where $\vec{k}_j$ is the wave vector of the respective Gabor wavelet φ_j and $\vec{d}$ is an estimated displacement that compensates for the rapid phase shifts (Wiskott et al., 1997).

2.7.2 Eigenfaces and Fisherfaces

When applying the linear subspace techniques, i.e. PCA and LDA, for face recognition, 2D face images are usually converted to a 1D feature vector by concatenating their rows or columns. Once the projection bases are learned from a set of training faces, they can be converted back to 2D images. These base images are thus called Eigenfaces and Fisherfaces for PCA and LDA, respectively.

2.7.3 The Difference Space

Canonical algorithms treat face recognition as a multi-class problem, i.e. each individual is a class. Some researchers also proposed the difference space to simplify face recognition to a two class problem. Such representation models the dissimilarities between faces. Let $T = \{I_1, \cdots, I_M\}$ be a training set of faces of K individuals, with several images of each of the subject. Two classes can be generated from T. The first is the intra-personal differences set, which are the dissimilarities in facial images of the same person:

$$CI = \left\{ I_p - I_q \middle| I_p \sim I_q \right\}$$

The second is the extra-personal difference set, which are the dissimilarities among images of different person in the training set:

$$CE = \left\{ I_p - I_q \middle| I_p \nsim I_q \right\}$$

The two sets thus define the difference space where face recognition can be represented as a two class problem.

2.8 Summary

A number of mathematical techniques have been introduced in this chapter, which will be applied in the following processes in the thesis:

Feature extraction: the mathematical origins of Gabor wavelets show that they are very powerful tools when applied to measure local spatial frequency and image structure. As a special wavelet, the Gabor wavelet analyzes images with the optimal spatial and frequency resolution. Motivated by the similarity of the 2D Gabor wavelet and the receptive field of the simple cells of the mammalian visual system, the wavelet family will be applied to extract local features from face images for recognition. Such local features will be robust against distortions caused by various expression, pose and illumination changes.

Feature enhancement and classification: once the robust feature set has been extracted by Gabor wavelets, a number of enhancement tools and classifiers can be further applied. While linear subspace techniques such as PCA and LDA have been shown to be able to enhance class separability, this chapter also gives theoretical evidence of further advantages of kernel methods. A number of techniques based on such methods (e.g. KPCA, GDA and SVM) have been introduced in this chapter and will be applied to enhance extracted Gabor features for recognition or classification in the following chapters.

Feature selection: both the mutual information and AdaBoost algorithms introduced in this chapter will be applied to select the most discriminant Gabor features for face recognition.

Chapter 3 Literature Review

This chapter gives a literature survey on state of the art face recognition algorithms, both 2D based and 3D based approaches are included. Particularly, the major concern of the thesis, i.e., Gabor wavelet based methods, are explored in detail.

3.1 2D Face Recognition Methods

Various approaches for 2D face recognition have been proposed in the literature, which can be classified into three categories: analytic (feature based), holistic (global) and hybrid methods. While analytic approaches compare the salient facial features or components detected from the face, holistic approaches make use of the information derived from the whole face pattern. By combining both local and global features, hybrid methods attempt to produce a more complete representation of facial images. Literature surveys on face recognition approaches can be found in (Chellapa, Wilson, & Sirohey, 1995) and (Zhao et al., 2000).

3.1.1 Analytic Methods

For analytic approaches, distances and angles between feature points on the face, shapes of facial features, or local features, e.g. intensity values extracted from facial features or components are usually applied for face recognition. The main advantage of analytic approaches is to allow a flexible deformation at the key feature points so that pose changes can be compensated. In (Brunelli & Poggio, 1993), both template and geometrical feature based analytic methods are implemented and compared. For template based method, facial regions are matched with templates of eyes, nose and mouth respectively and the similarity scores of each facial feature are simply added into a global score for face recognition. For geometrical feature based methods, eyes, mouth and nose facial features are firstly detected. The nose width and length, mouth position and chin shape features are then input to a Bayes classifier for identification. Figure 3-1 shows how these geometry features are measured, e.g. the chin shape is represented by the distance between the edge of the chin and the centre of the mouth. However, the experimental results favour the template matching approach.

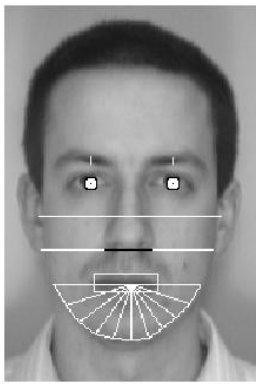


Figure 3-1 Geometric features used for face recognition (Brunelli et al., 1993)

A graph structure, called Dynamic Link Architecture (DLA), is proposed by Lades et al. (1993) to represent face images in. In this system, an elastic graph matching process is used to learn the representing graph of face images. Once faces are represented by appropriate graphs, Gabor features extracted from graph nodes, named Gabor jets, are then used for face recognition. Figure 3-2 shows two example face images overlaid with the representative graph (Lades et al., 1993). Later on, Wiskott et al (1997) extend DLA to Elastic Bunch Graph Matching (EBGM), where graph nodes are located at a number of selected facial landmarks. The EBGM has shown very competitive performance and been ranked as the top method in the FERET evaluation (Phillips et al., 2000). Details of Gabor wavelet based methods will be presented in section 2.

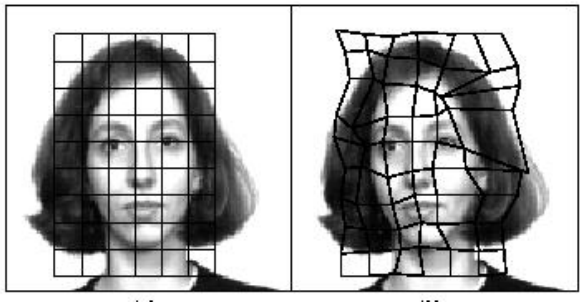


Figure 3-2 Face images represented by graphs (Lades et al., 1993)

The Hidden Markov Model (HMM), widely used to learn the state and transitional probabilities between a number of hidden states, has also been applied to face recognition. HMMs are normally trained from examples that are represented by a sequence of observations. The parameters of the HMM are firstly initialized and then adjusted to maximize the probability of the observation of the given training samples. The observation of test samples can then be input to the trained HMMs for classification according to the output probabilities given different HMMs. Samaria and Young (1994) first proposed a HMM architecture for face recognition. A face pattern is divided into several regions such as forehead, eyes, nose, mouth and chin. These regions occur in the natural order from top to bottom and they are used to form the hidden states of 1D or pseudo 2D HMMs. To train a HMM, each face image is represented by a sequence of observation vectors, which are constructed from the pixels of a sub window. Nefian and Hayes (1999) proposed the embedded 2D HMM, which consists of a set of super states with each super state being associated with a set of embedded states. Super states represent primary facial regions whilst embedded states within each super state describe in more detail the facial regions. As shown in Figure 3-3, transitions between embedded states in different super states are not allowed. Instead of using pixel intensities directly, the Discrete Cosine Transform (DCT) coefficients are used to form the observation vectors. Compared to 1D and pseudo 2D HMM, the system can perform more efficiently. Based on this work, Bai and Shen replaced DCT with the Discrete Wavelet Transform (DWT) for observation vector extraction (Bai & Shen, 2003a), the results show the performance improvement achieved. However, HMM based systems require lots of images for training, and are only capable of operating on small databases. The performance drops dramatically as the size of database is scaled up. As observed in our experiments, the accuracy of

Nefian and Hayes's method drops from 97.5% to 32.5% when the number of subjects rises from 40 to 200.

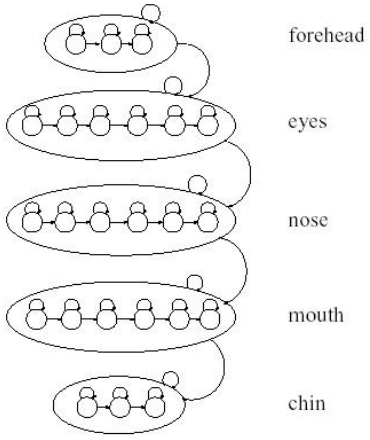


Figure 3-3 2D embedded HMM structure (Nefian et al., 1999)

As a hyper plane classifier, Support Vector Machines (SVM) have also been successfully applied to face recognition. A set of SVM classifiers is applied to extract different facial components and the grey values of each component are then combined into a single feature vector (Heisele, Ho, & Poggio, 2001). The component based method has been compared with a SVM classification based global method and the results show its robustness against variance of pose and illumination. However, the database consists of images from 5 subjects only and a large number of images are required to train those SVMs. Their later work (Huang & Heisele, 2003) used a 3D morphable model to generate synthesized images with different illumination and pose for training. As a result, only 3 training images of each person are required. However, the results are based on a database from 6 persons only. How the performance scales with the number of subjects in the database remains unknown.

3.1.2 Holistic Methods

Based on principal components analysis (PCA), Kirby and Sirovich (1990) first developed the well known Eigenface method for both face representation and

recognition. In this method, the whole face pattern is transformed to a feature vector and a set of training samples are used to compute Eigenfaces (Turk et al., 1991). PCA can achieve the optimal representation in the sense of maximizing the overall data variance. However, the difference between faces from the same person due to illumination and pose (within-class scatter) seems to be larger than that due to facial identity (between-class scatter). Based on this observation, Linear discriminant analysis (LDA) is applied for Fisher face methods (Belhumeur, Hespanha, & Kriegman, 1997). LDA defines a projection that makes the within-class scatter small and the between-class scatter large. This projection has shown to be able to improve classification performance over PCA. However, it requires a large training sample set for good generalization, which is usually not available for face recognition applications. To address such Small Sample Size (SSS) problems, Zhao et al (1998) perform PCA to reduce feature dimension before LDA projection, see Figure 3-4 for the different bases of LDA, PCA + LDA, and PCA projection. By using higher order statistical analysis, Independent Component Analysis (ICA) was first adopted by (Bartlett, Movellan, & Sejnowski, 2002) for face recognition, the work showed that ICA outperformed PCA. However, other researchers (Draper, Baek, Bartlett, & Beveridge, 2003) observed that when the right distance metric is used, PCA significantly outperforms ICA on the FERET database. Recently, kernel methods have been successfully applied to solve pattern recognition problems because of their capacity to handle nonlinear data. By mapping sample data to a higher dimensional feature space, effectively a nonlinear problem defined in the original image space is turned into a linear problem in the feature space (Scholkopf et al., 1999). PCA or LDA can subsequently be performed in the feature space and are thus called Kernel Principal Component Analysis (KPCA) and Generalized Discriminant Analysis (GDA) (Baudat et al., 2000). Experiments show that

KPCA and GDA are able to extract nonlinear features and thus provide better recognition rates in applications such as face recognition (Kim et al., 2002; Yang, Frangi, & Yang, 2004; Shen & Bai, 2004b).

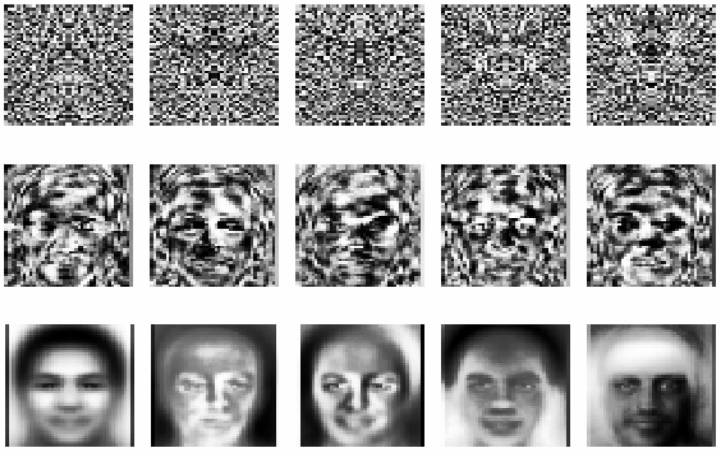


Figure 3-4 Different bases of linear projections: LDA, PCA + LDA and PCA bases are shown on the first, second and third row respectively (Zhao et al., 1998)

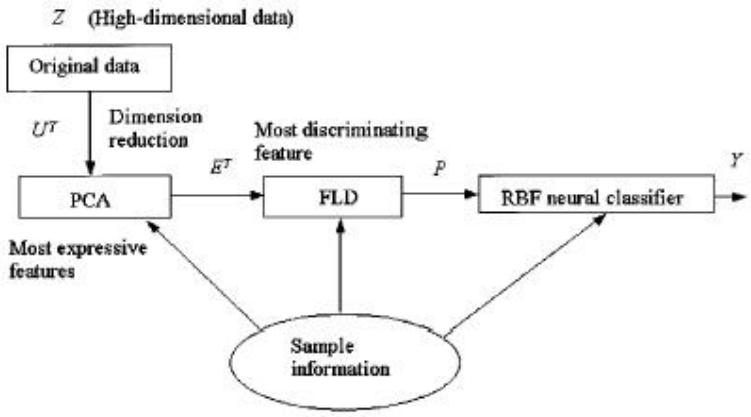


Figure 3-5 The diagram for a RBF based face recognition system (Er, Wu, Lu, & Toh, 2002)

Neural networks (Fleming & Conttrell, 1990; Er et al., 2002; Liu, 2004b) have also been used to classify global facial features. When face images were treated as 1D signals and wavelet analysis was used for feature extraction (Liu, 2004b), the Radial Basis Function (RBF) network was applied to the projection of face images to Fisherfaces for classification (Er et al., 2002). The diagram for Er's method is plotted in

Figure 3-5. While PCA + LDA were first used to decrease the feature dimension of face patterns, sample information was adopted to determine the structure and initial parameters of the RBF network.

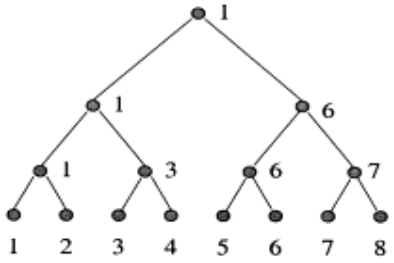


Figure 3-6 Binary SVM tree (Guo, Li, & Chan, 2001)

Since SVM is a binary classifier, (Phillips, 1999) turned the face recognition problem into a two class problem by introducing the difference space. Two classes, the dissimilarities between faces of the same person and dissimilarities between faces of different people, are designed in the difference space. A single SVM is trained to classify the intra-person and inter-person difference classes. The results on a difficult image set from the FERET database showed that SVMs outperformed the Eigenface method significantly. A binary tree system was adopted by (Guo et al., 2001) to use SVMs for the multi-class face recognition problem. The results on the ORL database and a larger face collection from several databases showed that SVMs achieve higher accuracy than Eigenface approach. In (Jonsson et al., 2002) each person is associated with a SVM that was trained to discriminate the face images from the same people and those from others. Both PCA and LDA were used for feature extraction and tested on a verification application. By applying different illumination normalization techniques, the results show that SVMs are robust and relatively insensitive to the feature space and pre-processing methods. However, when the representation feature already captures and emphasises the discriminatory information, e.g., features extracted using LDA or SVMs

lose their superiority in comparison with the simplest Euclidean distance + nearest neighbour classifier.

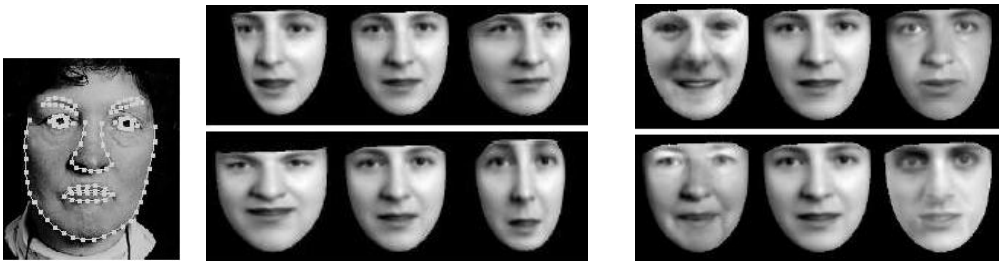
Global techniques work well for frontal view face images, but they are sensitive to translation, rotation and pose changes (Heisele et al., 2001). Usually normalization is an important and inevitable process for these methods. A small number of prominent points in the face such as eyes, nostrils or centre of the mouth are required to resize and rotate the input face image. After normalization, the input face image can be aligned with the model face and recognition can be performed thereafter.

3.1.3 Hybrid Methods

Hybrid methods utilize both local and global features for recognition. One of the early works is Pentland's modular Eigenfaces (Pentland, Moghaddam, & Starner, 1994). In this work, the eigenface technique is extended to the description and encoding of facial features, yielding eigenfeatures such as eigeneyes, eigennoses and eigenmouths. The experimental results show that the eigenfeatures outperform the eigenface method, the performance was further improved by using the combined representation of eigenfeatures and eigenfaces.

Another famous work is the Active Shape Model (ASM) and Active Appearance Model (AAM) proposed by (Lanitis, Taylor, & Cootes, 1997). In this work, Cootes' group use ASM and AAM to model the variance of shape and appearance respectively. Both ASM and AAM are learned from a large number of training images, which are then used to model test images. To recognize a face image, both ASM and AAM are adjusted to fit the new image, which generates a number of shape and texture parameters. Those parameters, together with the local profiles at model points, are used for face recognition. When 300 images (10 images per individual) are used as training images, the method achieves 92% accuracy for 300 test images. Figure 3-7 shows the landmarks

used to train the ASM, and the effects of varying the first two parameters of shape and appearance models.



(a) (b) (c)

Figure 3-7 Landmarks of ASM (a); variance of the facial shape (b); and appearance (c) (Lanitis et al., 1997)

3.2 Gabor Wavelet Based 2D Methods

Despite remarkable progress so far, the general task of face recognition remains a challenging problem due to complex distortions caused by various variations in illumination, facial expression and pose. It is widely believed that local features in face images are more robust against such distortions and a spatial-frequency analysis is often desirable to extract such features (Zhao et al., 2000; Scholkopf et al., 1997). With good characteristics of space-frequency localization, wavelet analysis seems to be the right choice for this purpose (Qian et al., 1996; Daubechies, 1990). In particular, among various wavelet bases Gabor functions provide the optimized resolution in both the spatial and frequency domains (Gabor, 1946; Daugman, 1985).

The Gabor wavelet was originally contributed by Gabor (1946) when he proposed to represent signals as a combination of elementary functions. The 2D counterpart of the Gabor elementary function was then introduced by (Granlund, 1978). Daugman (1985) reviewed the 2D Gabor wavelet family and presented evidence that the family can well model the 2D receptive-field profiles of simple cells in the mammalian visual cortex, and thus such visual neurons could optimize the general uncertainty relations for resolution in space, spatial frequency and orientation. From an information theoretic

viewpoint, (Okajima, 1998) derived the Gabor function as solutions for a certain mutual-information maximization problem. The work shows that the Gabor-type receptive field can extract the maximum information from local image regions. Due to the useful characteristics of Gabor functions, they have been widely and successfully applied for texture segmentation (Jain & Farrokhnia, 1991; Weldon et al., 1996), handwritten numerals recognition (Hamamoto et al., 1998), fingerprint recognition (Lee & Wang, 1999) and face recognition (Lades et al., 1993; Shen et al., 2004b; Wiskott et al., 1997; Liu et al., 2002). The wide application of Gabor functions has also resulted in different terminologies, which may be quite confusing for researchers. Some examples are Gabor wavelet, Gabor filter, Gabor expansion, Gaobr transform and Gabor function etc. Based on the fact that this study starts from joint time frequency analysis of signals, the terminology of Gabor wavelet is used in this thesis. While Gabor features are used to represent the features extracted by a set of Gabor wavelets, they are usually called jets when the wavelet family is applied at a certain facial feature point. A detailed survey on Gabor wavelet based face recognition methods, both analytic and holistic, will follow in the next section.

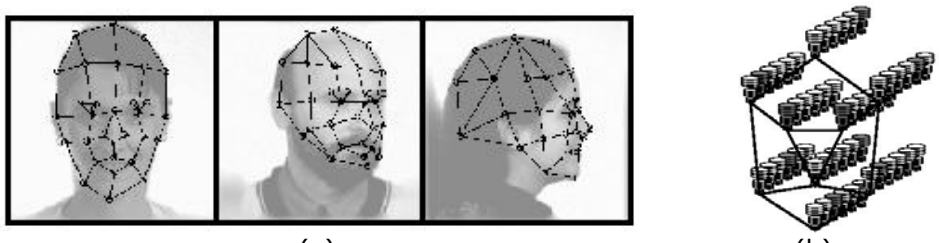
3.2.1 Analytic Methods

Analytic methods utilize the Gabor features, named Gabor jets, extracted from pre-defined feature points, on the face images for recognition. Different approaches mainly vary in the way to locate feature points for Gabor jet extraction, which can be classified into two categories: elastic graph matching based methods and non graph matching based methods. For elastic graph based analytic methods, a graph is first placed at an initial location and deformed using jets to optimize its similarity with a model graph. Non-graph based methods locate feature points manually or by colour or edge etc.

information. Once the location process is completed, recognition can then be performed using Gabor jets extracted from those feature points.

3.2.1.1 Elastic Graph Matching Based Feature Points Location

Dynamic Link Architecture (DLA) (Lades et al., 1993) and Elastic Bunch Graph Matching (EBGM) (Wiskott et al., 1997) are two famous Gabor jet based methods using elastic graph matching for face representation. Graph matching based methods normally require two stages to build the representing graph g^I for a face image I. During the 1st stage, a model graph g^M is shifted within the input image while keeping its form rigid. The rigid graph is initialized at an arbitrary position in the input image. A cost function $S(g^I, g^M)$ is defined (see Eq. 3.2) and the position is updated until a minimum value of the function is reached. The global move procedure is then followed by individual vertices diffusion during the 2nd stage. The vertices of the model graph are visited in a random order and are shifted by a random vector $\vec{d}$ within a topological constraint $\vec{T}$ to encode the local distortions due to rotations in depth or expression variations. It is actually the deformation of the vertices that makes the graph matching processing elastic.



(a) (b)

Figure 3-8 Face adapted graphs for different poses (a) and an example face bunch graph (b) (Wiskott et al., 1997)

In DLA (Lades et al., 1993), a model graph is built for each individual face in the gallery and the graph matching process is required to learn the representing graph for a new face image. The model graph in DLA is a rectangular graph, with each node

labeled by Gabor jets. Two sample face images with overlaid representation graphs are shown in Figure 3-2. The graph shown in (b) is built by applying the 2 stage graph matching process using (a) as the model graph. Based on DLA, Wiskott et al. (1997) further developed a more appropriate graph structure, called EBGM, to represent faces. Compared with the rectangle graph used in (Lades et al., 1993), the new method employs object adapted graphs and each node refers to specific facial landmark. Figure 3-8a shows the adapted graph grids for faces with different poses, one can observe that such structure is more suitable for face images. Since matching with each individual model graph is very computationally expensive for large galleries, they also developed a technique called the Face Bunch Graph (FBG, shown in Figure 3-8b) to avoid such a process. A bunch is a set of jets taken from the same node from different model graphs. This requires a set of aligned model graphs, such that a given node always refers to the same facial features. 80 manually built model graphs are used in (Wiskott et al., 1997) to build the FBG, which is then used as the only model graph to build the representing graph for an input face image using the 2 stage graph matching process.

Since the representing graph of a face image is normally associated with a set of corresponding Gabor jets, jet similarity plays a very important role in the definition of the cost function $S(g^I, g^M)$ to match two graphs. Two different functions can be used to compare jets (Wiskott et al., 1997). The first one, $S_m(J, J')$, using magnitude information only, generates more smooth output when a fixed $J'(\vec{z})$ is compared with jets $J(\vec{z}+\vec{d})$ located at varied positions with displacement $\vec{d}$. The other one, $S_p(J, J')$, takes phase into consideration, is more sensitive to displacements and potentially more discriminative since jets with the same magnitudes but different phase can be distinguished. For a labeled graph with nodes $\{\vec{z}_1 \vec{z}_2 ... \vec{z}_N\}$ and edges $\Delta\vec{z}_e = \vec{z}_i - \vec{z}_j$,

$e = 1,2,...,E, \quad i = 1,2,...N, j = 1,2,...,N$, the similarity of a model graph g^M and a variable graph g^I is evaluated by a cost function in DLA as (Lades et al., 1993):

$$S(g^I, g^M) = \lambda \sum_e (\Delta \vec{x}_e^I - \Delta \vec{x}_e^M)^2 - \sum_n S_m(J_n^I, J_n^M) \tag{3.1}$$

where λ determines the relative importance of jet similarities and the topography term. J_n is the jet at node *n* and $\Delta \vec{x}_e$ is the distance vector to label edge *e*. This function does not take the phase of jets into consideration. Similarly, the quality of matching between an image graph *I* and the FBG *B* is evaluated by (Wiskott et al., 1997):

$$S(g^I, g^M) = \frac{\lambda}{E} \sum_e \frac{(\Delta \vec{x}_e^I - \Delta \vec{x}_e^B)^2}{(\Delta \vec{x}_e^B)^2} - \frac{1}{N} \sum_n \max_m S_p(J_n^I, J_n^{B_m}) \tag{3.2}$$

where B_m denotes the *mth* model graph of the bunch graph B. The cost functions thus defined takes the similarity of both jets and graph geometry into consideration. Other definitions of the cost function can also be found in (Rong, Su, & Lin, 2002). In the 2nd stage of graph matching, the graph nodes are also shifted within a topographical constraint $\vec{T}$ to model the local face distortions. Wiskott (1999) used a simple rectangular graph model to investigate the role of topographical constraints for face recognition. The primitive graph models with different strengths of topographical constraints are compared with a more sophisticated system using bunch graphs. The results show that the constraints are quite useful when the variations in illumination, scale and background are small. His work also compared different jet similarity measure functions and the results suggest that the function with phase yields better matching results than the one without phase when drastically changing illumination is not available.

Based on the elastic graph matching framework, a number of varitions have been proposed in the literature. Mu and Hassoun (2003) proposed a group shift/deformation algorithm. The algorithm clustered rectangular graph nodes into groups (eyes, mouth

and nose etc.) according to their locations. All the graph nodes in the same group move together in the rigid matching stage, while local deformation is allowed in the 2nd step, see Figure 3-9 for details. The results on two databases show that the proposed group shift algorithm achieved better performance than the standard elastic graph matching algorithm. Elastic graph matching has also been applied to face authentication by Duc et al. (Duc, Fischer, & Bigun, 1999). The importance of the rectangular graph nodes is measured by a criterion specially designed for acceptance and rejection of the candidate. The criterion is small when the candidate is the right person, and large in case of an impostor. The Fisher discrimination criterion turns out to be the right one. They show that a feature consisting of only Gabor jets extracted from those important nodes not only reduces the feature dimension, but also improves the recognition performance significantly. Since the elastic graph matching process is very computationally expensive, they also tested the significance of the elastic steps by simply dropping them, which is equivalent to setting $\lambda = \infty$ in the graph similarity function. The comparison of performance obtained with and without the deform step shows that the elastic matching slightly increases the performance, but has less influence than weighting of the graph nodes.

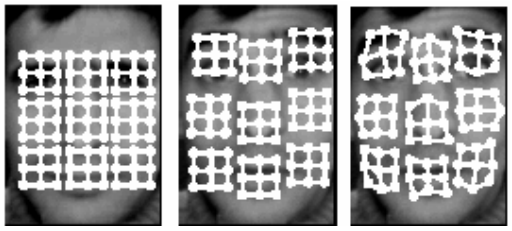
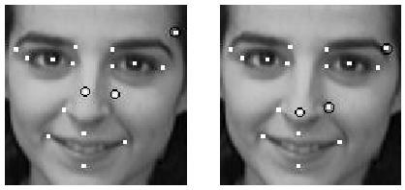


Figure 3-9 The group shifting/deformation algorithm (Mu et al., 2003)

Liao and Li (2000) reduced the nodes of the bunch graph to only 17 facial feature points, all of which have clear meanings and exact positions. A collection of 70 face images with manual marks at correct facial feature points are used to construct the bunch graph.

Once the FBG is determined, the facial feature points can be detected automatically for the new input image by the elastic graph matching process. Since some feature points may be located at the wrong places, a graph adjusting stage is proposed to correct the wrongly positioned points. Figure 3-10 shows the results of automatic facial feature point detection. The three misplaced feature points marked by black circles (Figure 3-10a) are corrected by the graph adjusting process (Figure 3-10b). Instead of using the rigid matching step, Jiao et al. (Jiao, Gao, Chen, Cui, & Shan, 2002) used face structure knowledge and grey intensity information to locate the facial features, e.g. eyes and mouth. Once the features are located, the position of the bunch graph is initialized and the elastic deformation step is then used for feature position refining and adjusting.



(a) (b)

Figure 3-10 17 facial feature points and the results of graph adjustng (Liao et al., 2000)

3.2.1.2 Non-graph Matching Based Feature Points Location

Due to the computational complexity of the elastic graph matching process, a number of works have also proposed other techniques for feature point location. Some works locate the feature points manually (Escobar & Ruiz-del-Solar, 2002; Gokberk, Irfanoglu, Akarun, & Alpaydin, 2003; Wang & Qi, 2002; Chung, Kee, & Kim, 1999) and Gabor jets extracted at those points are then subjected to a sophisticated classification system for recognition. Escobar et al (2002) proposes to use Log-Polar images for Gabor feature extraction. The face image is Log-Polar transformed before it is convolved with Gabor wavelets. This technique is supposed to be more robust against the variance of

scale and rotation. In this system, facial feature points are located manually and the coordinates are Log-Polar transformed as well.

Wu et al. (Wu, Yoshida Y., & Shioyama, 2002) used both colour and edge information to extract facial organ regions, feature points are then detected by applying the SUSAN corner detector. 12 Gabor wavelets with tuned parameters are designed and used for both feature point location and feature extraction. Face structure and intensity were also used by Jiao et al. (2002) to locate facial features, e.g. eyes and mouth.

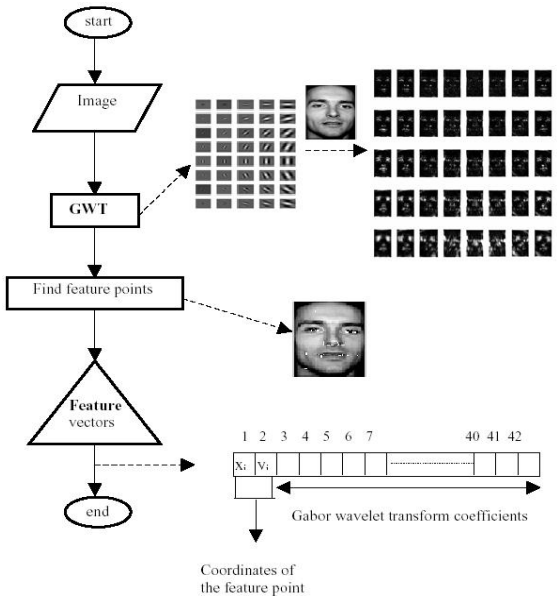


Figure 3-11 Flowchart of variable feature points location (Kepenekci, 2001)

Instead of using the pre-defined facial features such as eyes, nose and mouth, some researchers have proposed to locate feature points in the face images which contain interesting information (Kepenekci, 2001; Hjelmas, 2000). These points are not necessarily specific feature points, but they are usually positioned around facial features. Hjelmas applied a family of 24 Gabor wavelets to the face image and the magnitudes of the convolution results at each location in the image are summed to result in the filtered image. The centre area of the face is emphasized by Gaussian weighting and a maxima

selecting algorithm is used to locate the feature points with useful information. Similar to the method in (Hjelmas, 2000), points with high-energized Gabor wavelet response are found by searching the pixels in a sliding window (Kepenekci, 2001). 40 Gabor wavelets are convolved with the face image and the searching process is applied to each of the 40 resultant images. The number of feature points and their locations vary for different face images.

3.2.1.3 Face Similarity Measures and Recognition

Once a face has been represented by a set of Gabor jets extracted from located feature points, face recognition is a trivial step. For graph matching based methods, the identity of a test image is determined by the statistics of graph similarity values between test graphs and all model graphs in (Lades et al., 1993). The similarity function of two facial images is simply an average over the similarities between pairs of corresponding jets in (Wiskott et al., 1997) and (Liao et al., 2000). After comparing two strategies for combining local jet similarities, (Mu et al., 2003) suggests that a voting strategy should be used. The set of Gabor jets extracted at different feature points could also be combined into a long feature vector and a simple distance measure could be applied for classification (Duc et al., 1999; Wang et al., 2002; Jiao et al., 2002). Three different distance measures are tested in (Jiao et al., 2002) and the results suggest that the city block distance metric achieves better performance than cosine methods. More sophisticated classifiers have also been applied to the combined feature vector for recognition, e.g., a Bayesian classifier is adopted in (Wang & Tang, 2003) and improvements have been achieved over the system using direct correlation of Gabor features for classification. Chung et al. (1999) applied PCA to the extracted Gabor response at predefined facial feature points such that local variations can be included to overcome the shortcoming of PCA.

Both methods, whether the feature points are located by edge detector (Wu et al., 2002) or manually (Escobar et al., 2002), use the average of the jet's similarity as a measure of face graphs. The jet similarity function without taking phase into account is used. Since the correspondences of jets between two facial images are unknown, only jets with similarity above a preset threshold are taken into consideration (Kepenekci, 2001). The image similarity of two facial images is calculated as the mean of the similarities of the selected jets. To include information of topological similarity, the number of similar jets could also be taken into the similarity function. In this case, the overall similarity of a test image and a reference image is a weighted sum of the image similarity and the number of similar jets.

3.2.2 Holistic Methods

While analytic methods utilize the Gabor jets extracted from prominent feature points for recognition, holistic methods normally extract features from the whole face image. An augmented Gabor feature vector (Liu et al., 2002) can be derived by concatenating the Gabor jets at all pixel locations. Since the feature vector consists of all useful information extracted from different frequencies, orientations and locations, this representation can produce discriminant features for recognition. Similar to typical holistic face recognition methods, faces need to be detected and normalized in size and orientation prior to recognition. Various works have shown that such Gabor features are much more robust than grey-level intensity values against the mis-alignment caused by the normalization procedure (Shan, Gao, Chang, Cao, & Yang, 2004).

A number of researchers have developed different recognition systems based on this feature vector. In Liu's early work (Liu et al., 2002), he applied the Enhanced Fisher linear discriminant Model (EFM) on the Gabor feature vector for face recognition, results show that the novel Gabor-Fisher Classifier outperformed both PCA and LDA.

Since the 40 Gabor filtered images are concatenated together to form a feature vector (see Figure 3-12), the dimension is huge, e.g., 163,840 for images with size 64×64. As a result, downsampling is first used to reduce the dimension to manageable size. He also applied Independent Component Analysis (ICA) (Liu & Wechsler, 2003) on the augmented feature vector and developed a so-called Independent Gabor Feature (IGF) for recognition. The results show that ICA performs significantly better than eigenfaces. One of his recent work (Liu, 2004a) utilized Kernel PCA with fractional power polynomial kernel to reduce the dimension of the extracted Gabor feature vector and enhance the discriminative power at the same time. However, no direct comparison among those proposed approaches is presented. Shen and Bai (Shen et al., 2004b; Shen & Bai, 2004c) mapped the augmented Gabor features to kernel space, i.e., the extracted Gabor feature is analyzed by Generalized Discriminant Analysis (GDA), or Kernel Direct Discriminant Analysis (KDDA) for further feature enhancement. Experimental results show that kernel methods achieve much better results than linear methods such as PCA and LDA. The work of both Liu (Liu et al., 2002) and Shen (Shen et al., 2004b) have shown that Gabor feature based methods can achieve significant improvement over those using raw pixels, which proved the discrimination ability of Gabor feature. Similar work can also be found in (Fan, Wang, Liu, & Tan, 2004), which applies Null LDA (NLDA) to the augmented Gabor feature vector for recognition.

Once the dimension of extracted feature vector has been reduced and discrimination ability enhanced by a certain subspace analysis, simple nearest neighbour classifier and Euclidean distance measure can be applied for classification. When the simple Euclidean distance measure seems to be enough, research results do suggest that different distance measures may affect the performance of system and an appropriate distance measure has to be chosen for different subspace analysis approaches (Liu et al.,

2002; Shen et al., 2004b). More complex classifiers, e.g. Support Vector Machine (Chi, Dai, & Zhang, 2004) and Nearest Feature Space (Zhu, Vai, & Mak, 2004), could also be applied to the enhanced features for possible improvement of accuracy. However, such kinds of system are more complex and the improvement is not guaranteed.

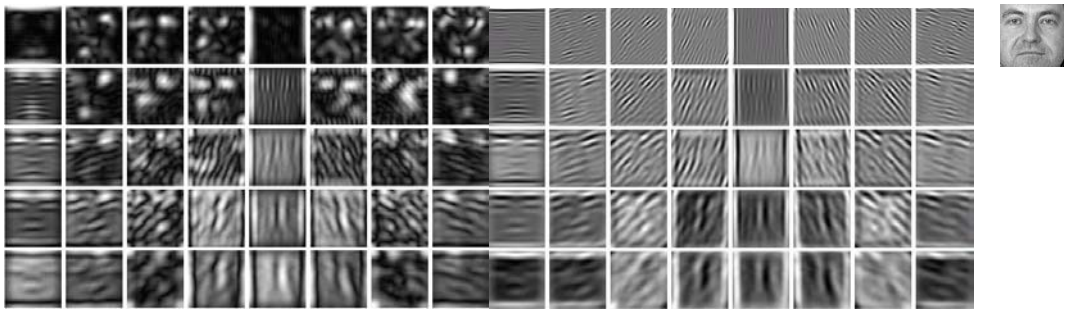


Figure 3-12 Convolution results of a face image with 40 Gabor wavelets

A quite different method proposed by Ayinde and Yang uses rank correlation of Gabor-filtered images for face recognition (Ayinde & Yang, 2002). Instead of concatenating all of the filtered images together, their method compares the filtered images separately. Three Gabor filtered images with selected orientation and kernels, together with the original face image and the neighborhood averaging of two filtered images, are used to represent the faces. Rank correlation values derived from the six representing images are then weighted together to yield the overall matching score of two face images. A face is matched to the subject that produces the highest similarity score computed from the six rank correlation values. Since the weighting parameters need to be decided from the training images, the optimisation process is very length. It is reported in this paper that the process takes 35 minutes to complete a run for parameter determination using 200 training images.

3.2.3 Gabor Wavelet Network

Whilst most of the works in the literature use Gabor wavelets for feature extraction, the characteristics and compression ability of wavelets have not been fully explored. Reconstruction of the signal from compressed wavelet coefficients is actually one of the main reasons that lead to the wide application of wavelets in the real world (Strang & Nguyen, 1996; Mallet & Zhong, 1992). Due to the nonorthogonality of Gabor wavelets, application of Gabor wavelets in signal reconstruction is very limited. Credits must be given to Krueger, who proposed the use of Gabor Wavelet Networks (GWN) for object representation and face processing (Kruger et al., 2000; Kruger & Sommer, 2002b; Kruger et al., 2002a). Originating from the idea of wavelet networks (Zhang & Benveniste, 1992) and the fact that Gabor functions have been widely applied to feature extraction, Krueger proposed to use a set of Gabor wavelets $\mathbf{\Psi} = (\varphi_1 \varphi_2, \cdots \varphi_N)^T$ with associated weights $\mathbf{W} = (w_1 w_2, \cdots w_N)^T$ to represent a face image. The set of Gabor wavelets and weights are obtained through optimizing the objective functional of reconstruction error $E = \min \left\| I - \sum_i w_i \varphi_i \right\|^2$. The two vectors $\mathbf{\Psi} = (\varphi_1 \varphi_2, \cdots \varphi_N)^T$ and $\mathbf{W} = (w_1 w_2, \cdots w_N)^T$ now define the GWN for representing image I. Given the optimal GWN of an image I, it can be reconstructed by a linear combination of the weighted wavelets: $I = \sum_i w_i \varphi_i = \mathbf{W}^T \mathbf{\Psi}$. The quality of the reconstruction of course depends on the number of wavelets used and can be varied to reach desired precision. Figure 3-13 shows the images reconstructed with 16, 52, 116 and 216 Gabor wavelets (left to right).

Figure 3-13 Original image and the reconstructed image with different number of wavelets (Kruger et al., 2002a)

Since Gabor wavelets are nonorthogonal bases, linear projections of a new pattern on them do not produce the correct weights. As a result, dual Gabor wavelets $\widetilde{\mathbf{\Psi}} = (\widetilde{\varphi}_1 \widetilde{\varphi}_2, \cdots \widetilde{\varphi}_N)^T$ have to be found to compute the weights:

$$\mathbf{W} = I\,\widetilde{\mathbf{\Psi}} \;\; with \;\; \widetilde{\mathbf{\Psi}} = \left(\mathbf{\Psi}^T \mathbf{\Psi}\right)^{-1} \mathbf{\Psi}^T \tag{3.3}$$

Once the GWN is learned to represent a face image, the representations can be used for recognition. Since the number of wavelets and weights may vary for different images, a special distance measure has been designed in (Kruger et al., 2002b) for similarity measurement. Recently, Zhang et al. (Zhang, Zhang, Huang, & Tian, 2005) proposed the concept of the Subject Dependent Gabor Wavelet Network (SDGWN), which is learned from all of the training images of the same subject. Instead of representing each subject image with different GWNs, their method uses the same GWN model to represent all images from the same individual. The SDGWN was then further combined with a recent proposed neural network model, named Kernel Associative Memory (KAM) for face recognition. The results on FERET, ORL and AR face databases show that this method achieved better performance than other popular approaches.

3.2.4 Performance Evaluation

With a lot of face databases available, evaluation of different face recognition algorithms is always one of the most difficult tasks. Even when the same database is used, different papers may use different parts of the database for experiments. Moreover, the partitioning of training images, gallery images and test images may also vary. For example, the results of (Wiskott et al., 1997) were reported using 250 fa and 250 fb images from FERET, while those from (Liu et al., 2002) were reported using 600 frontal FERET images. In (Lades et al., 1993), the database consists of images captured from 87 people. Subjects were asked to keep in standard pose, look 15^o to the right and make a random expression. The standard images are used as the model and the images

with different poses and expressions are used as two probe sets for testing. Accuracy of 98% was achieved for elastic bunch graph methods when frontal view faces were used for testing (Wiskott et al., 1997), where neutral frontal view faces (fa) were used as the model gallery and frontal view faces with different facial expression (fb) were used as probe images. When half profile faces or profile faces were matched with the frontal faces, the accuracy drops significantly. A number of databases are tested in (Zhang, Yan, & Lades, 1997) and the results of their algorithm are compared with state of the art algorithms such as Eigenface, elastic graph matching and neural network, etc. It is claimed that their algorithm is competitive to the popular methods and their algorithm achieves higher performance than most of other algorithms when the FERET database is concerned. Liu et al. (2002) use 600 FERET frontal face images with different illumination and facial expressions from 200 subjects for performance evaluation. The eyes of face images were manually detected and used to normalize the scale and rotation. Two images of each person were randomly chosen as training images while the remaining image was used for testing. 100% accuracy was achieved when the dimension of feature was set as 65. A pose estimation module is also developed in (Liu, 2004a) and the algorithm is tested using the CMU PIE database where faces with different poses are available. The accuracy of 96% reported in (Ayinde et al., 2002) was achieved when 9 images of each person in the ORL database were used for training. Table 3-1 summarizes the database and recognition rate of different Gabor feature based algorithms. All of the recognition rates listed in the table are for frontal view faces only, results for half profile and profile faces from FERET database can be found in (Wiskott et al., 1997) and (Hjelmas, 2000). Since the performance for methods group shifting /deformation (Mu et al., 2003) and weighted EBGM (Duc et al., 1999) are

reported with False Acceptance Rate (FAR) and False Rejection Rate (FRR), their results are not included in the table.

Algorithms			Test Database	Recognition Rate (%)
Local Methods	Elastic Graph Matching Based Methods	DLA (Lades et al., 1993)	Own	88
		EBGM (Wiskott et al., 1997)	FERET	98
		Liao and Li's method (Liao et al., 2000)	Yale	96.4
		Gabor + Bayesian (Wang et al., 2003)	XM2VTS	97.1
			AR	93.3
		Face structure based facial feature detection (Jiao et al., 2002)	Own	93.3
			ORL	94.5
	Non-Graph Matching Based Methods	Variable facial feature points (Kepenekci, 2001)	Stirling	100.0
			AR	100.0
			ORL	95.3
			FERET	96.3
		Edge/color based facial feature points detection (Wu et al., 2002)	Own	92.8
		Log-Polar + Gabor (Escobar et al., 2002)	Yale	88.9
Global Methods		Gabor-Fisher (Liu et al., 2002)	FERET	100.0
		Gabor – ICA (Liu et al., 2003)	FERET	98.5
			ORL	100
		Gabor – Kernel PCA (Liu, 2004a)	CMU PIE	95.3
			FERET	99.5
		Gabor-Kernel (Shen et al., 2004b)	FERET	97.5
			ORL	100
		Gabor + rank correlation (Ayinde et al., 2002)	ORL	96.0
			UMIST	97.5
Gabor Wavelet Networks		GWN (Kruger et al., 2002b)	Yale	97.8
			Mancheste r	93.3
		SDGWN + KAM (Zhang et al., 2005)	FERET	99.6
			ORL	100
			AR	96.5

Table 3-1 List of Gabor wavelet based face recognition algorithms and accuracy

A few works in literature have also compared Gabor feature based methods with other popular face recognition algorithms. A Dynamic Link Architecture based algorithm is evaluated as more robust than eigenface methods and neural network approaches (Zhang et al., 1997). A combination of four databases: MIT, ORL, Weizmann and Bern were used to evaluate different algorithms. While the performance of the eigenface

method deteriorates significantly as lighting variation increases, the elastic matching algorithm, on the other hand, is insensitive to lighting, face poses, and expression variations and therefore is more versatile. An accuracy of 93% was reported for the DLA algorithm, which is much higher than that of Eigenface methods (66%). Kalocsai et al. (Kalocsai, Zhao, & Elagin, 1998) attempt to compare the performance of machine face recognition systems with that of humans: 64 volunteers performed a sequential face matching task and their error rate and reaction time was recorded as the psychophysical data. Two face recognition models, the DLA and PCA-LDA models were also applied to the same image test set and the results were compared quantitatively and qualitatively. The analysis shows that both models are correlated to human performance, however, the DLA model seems to capture human performance better than PCA-LDA model.

Several large databases and evaluation protocols have also been available in literature such that different algorithms can be compared in the same framework. In 1996 and 1997, the FERET evaluation methodology and benchmark were designed to evaluate state of the art face recognition algorithms (Phillips et al., 2000). Different test sets were designed in the evaluation to test the robustness of face recognition methods against variance caused by various expressions, illuminations and capture times. A number of systems such as PCA, PCA + LDA, neural network and Bayesian methods were evaluated and the results show that EBGM achieved the top performance. To make the testing be as closely as possible with real authentication applications, the BANCA database (Baillere & Bengio, 2003; Messer et al., 2004) has also been released recently to replace the XM2VT database (Messer, Matas, Kittler, Luettin, & Maitre, 1999) for evaluation of face verification algorithms. Organized by the University of Surrey (UK), more than 10 research institutes participated in the face verification

competition (FVC2004), which was based on the BANCA database. Several protocols were designed in this competition to test the robustness of algorithms against variance of image quality, face poses and illumination. The results show that two methods using Gabor wavelets for feature extraction demonstrated the top performance (Messer et al., 2004). Based on the comparison of the Gabor feature based methods with state of the art algorithms and the results of FERET evaluation and FVC2004, we believe that Gabor wavelets might be the best choice to extract features for face recognition. The features could be extracted either locally or globally, and then different classification approaches can be applied.

3.2.5 Complexity of Gabor Feature Based Methods

Despite the advantages of Gabor wavelet based algorithms in recognizing face images with different illumination, pose and expression, they require high computational efforts. Even when a parallel computer system was used, it was reported in (Lades et al., 1993) that the convolution of a 128×128 pixel image with 40 Gabor wavelets took about 7 seconds. When 23 transputers were used, the comparison of an image to a stored face model took 2 to 5 seconds, while the identification of a probe face in a database of 87 people took about 25 seconds. For the elastic bunch graph matching algorithm, the location of face, detection of facial feature points and matching with FGB together take less than 30 seconds on a SPARC station 10-512 (Wiskott et al., 1997). Since fewer graph nodes are used and the similarity of graphs is simply an average over the similarities between pairs of corresponding jets, the comparison of an input face against a database of 250 people took less than 1 second. As a result, the main computational loads for graph matching based analytic methods are from the process of the convolution of the image with the family of Gabor wavelets, and the elastic graph matching step. Fast Fourier Transform (FFT) and Inverse Fast Fourier Transform (IFFT)

can be used to speed up the convolution process, i.e. both Gabor wavelets and the image are transformed to frequency domain using FFT and the product is then transformed back to spatial domain using IFFT. The whole convolution process can thus be completed within 2 seconds for images with size 128×128 on a Pentium 4 1.8G HZ PC. However, the 2 stage elastic graph matching process remains a time consuming step. A natural way is to replace part of, or the whole graph matching process with a faster method implementing similar function.

Jiao et al. (2002) replaced the rigid matching step with a structure knowledge and grey intensity information based facial features location process. Once the features are located, the position of the bunch graph is initialized and the elastic deformation step is then used for feature position refining and adjusting. However, the time saved compared to the standard elastic graph matching process is not reported. Duc et al. (1999) proposed a coarse to fine rigid graph matching method to speed up the 1st stage process, which is based on a Gaussian pyramid structure. They also tested the significance of the elastic step by simply dropping them, which results in 3% increase in classification error. The performance drop due to the elimination of the deform step is not significant and can be compensated by other enhancements, e.g. weighting of the graph nodes. The whole elastic graph matching process could also be replaced by a robust facial feature location process. In (Wu et al., 2002), once the image has been preprocessed using Gabor wavelets, facial feature points are detected using color, edge, Gabor features and corner information, which only takes about 0.15 seconds. When 12 purposely designed Gabor wavelets are applied for facial feature point extraction, it is reported that the processing time of the Gabor transformation takes about 3 seconds with a 533MHz Celeron processor. The processing time could have been further

reduced with a more powerful PC. However, the feature point location algorithm itself has to be robust against the variation of illumination, pose and expression.

Similarly, the computationally intensive convolution processes for Gabor feature based holistic methods could also be speed up by using FFT and IFFT. However, the dimension of the extracted Gabor feature is incredibly huge, e.g. 655,360 for an image with size 128×128 when 40 wavelets are used. Although downsampling could be used to reduce the feature dimension to a certain magnitude, the dimension after downsampling is still very high, e.g. 16,384 with a downsampling rate of 40 (Shen et al., 2004b). As a result, high memory capacity is required to save the features of face templates. In addition, both the training and application of classifiers using such high dimensional features would be very time consuming. The Gabor feature representation of a face image is substantially compressed when a GWN is used. 52 wavelets have been shown to be sufficient for real time pose estimation and face tracking (Kruger et al., 2002b). However, the GWN optimizing process given an image requires a high computational cost. It was reported in (Kruger et al., 2002b) that it takes about 30 seconds on a 750-MHz processor to optimize a GWN with 16 wavelets, even when a coarse-to-fine strategy has been adopted.

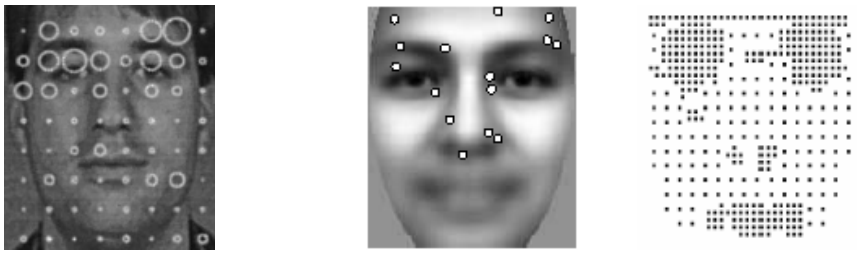
3.2.6 Optimization of Gabor Wavelets for Feature Extraction

As described in the last section, a number of methods have been proposed to reduce the computational complexity of Gabor feature extraction, e.g. FFT or using alternative facial feature location approaches, etc. Some researchers have also tried to optimize the Gabor representations by using a feature selection scheme. The dimension of Gabor features could thus be reduced and the feature will be more robust against the influence of noise. These optimization methods can be mainly classified into the following categories:

3.2.6.1 Optimization of Locations

A local linear discrimination criterion has been developed in (Duc et al., 1999) to measure the importance of different nodes on the rectangular graph representing face images. By using only the Gabor jets located at significant nodes, not only is the feature dimension reduced, but the classification performance is also improved. The discrimination criterion is similar to the Fisher measure (Fisher, 1936) such that the variance between samples of the same individual is minimized. Another interesting work models the feature location optimization objective as a subset selection problem (Gokberk et al., 2003). They tested three different Gabor jet representation schemes: a) rectangular graph with sparse nodes; b) face adapted graph with nodes located at prominent facial features only, e.g. eye corners, mouth corners, etc. c) the whole convolution result including all pixels in the image. Different feature selection methods such as best individual feature (BIF), sequential forward selection (SFS), sequential float forward search (SFFS) and genetic algorithm (GA) were tested and the results show that GA with representation scheme c) achieved the best performance. One can observe that most of the significant jets are located at the periphery of facial features. However, the results do suggest that the best locations to represent face images using Gabor jets may not necessarily be exactly at the facial features. PCA is performed on the augmented feature vectors in (Liu, Lam, & Shen, 2004). They argue that the summation of the eigenvectors at a particular position represents the corresponding variations among training images and thus reflects the corresponding importance in distinguishing human faces. Each pixel in the image is then classified as either a “key point” or “assistant point” based on this criterion. Different sampling intervals are adopted on the key and assistant points and a Gabor feature vector of lower dimension can thus be generated. LDA is finally applied to the resultant feature vector for face

recognition. See Figure 3-14 for the feature locations selected by different algorithms. As can be seen from the figure, most of the significant locations are around facial features, e.g. eyes, nose and mouth etc.



(a) (b) (c)

Figure 3-14 Significant locations selected by different algorithms: (a) a local discrimination criterion ranked jets location, significances are proportional to the radii of the circles; (b) the 15 most important locations selected by GA; (c)2×2 sampling for key points while 4×4 sampling for assistant points

3.2.6.2 Optimization of Gabor wavelets

Instead of optimizing the locations to extract jets for face representation, a few works have tried to optimize the Gabor wavelet basis used for feature extraction. Wang and Qi (Wang et al., 2002) applied GA to select the optimized Gabor wavelet basis for feature extraction. 34 easily identifiable landmarks, located manually on each image, are selected to represent faces. A set of Gabor wavelets with 4 scales and 6 orientations is then designed as candidates and the aim of the GA is to then select the optimal subset as a basis for face representation. To reduce the computation burden on the GA, they also proposed to use information complexity as a fitness measure of the chromosome. Face recognition is then performed based on the 4 optimal basis selected by GA and substantial improvements over the eigenface method have been observed.

In summary, most of the works available in literature either select locations where a fixed set of Gabor wavelets are applied, or optimize the wavelet basis to be convolved at a fixed set of feature points. Since different parts of natural objects usually display

various local characteristics, an improved method should apply the optimal wavelets at the most appropriate locations for feature extraction.

3.3 3D Face Recognition Methods

With most of the 2D recognition methods focusing on frontal view face images only, 3D models have been adopted to recognize faces with any pose. One of the representative works using a 3D model is described in (Romdhani, Blanz, & Vetter, 2002). This work performs face recognition in an analysis-by-synthesis fashion. The algorithm uses linear equations to recover the shape and texture parameters irrespective of pose and lighting conditions of the face image. Those parameters are then used for recognition. However, the model fitting process takes quite long time, e.g., 8 minutes on a Pentium III 800 MHz PC. Similar work can also be found in (Zhao & Chellapa, 2000; Lee & Ranganath, 2003). In these works, a 3D face model was usually used to synthesize images with different illumination and poses from a frontal face image, 2D techniques are then applied to the synthesized images for recognition.

With the development of 3D capture systems, face recognition using 3D facial data is also attracting much attention. (Beumier & Acheroy, 2000) developed both surface matching and central/lateral profiles for recognition, the results show that the two methods give the same level of performance. Other techniques used for 3D face recognition are Extended Gaussian Image (EGI) (Tanaka, Ikeda, & Chiaki, 1998) and point signature (Chua, Han, & Ho, 2000). Some works also applied 2D techniques to 3D range data for recognition, e.g., 3D Eigenfaces (Hesher & Erlebacher, 2002). In addition to using 3D data only, multi-modal 3D+2D face recognition has also been proposed (Wang, Chua, & Ho, 2002). In this work, Gabor wavelet responses in 2D and point signatures in 3D are integrated to an augmented vector for feature representation. Classification is done by SVMs.

Despite the overall optimism about 3D face data relative to 2D face images, it is pointed out by (Bowyer, Chang, & Flynn, 2004) that there are still significant limitations in current 3D sensor technology and most current 3D face recognition algorithms do not handle expression variations well. While 3D shape is defined independent of illumination, it is sensed dependent of illuminations. "Holes" may occur in areas where data is missing, even under ideal illuminations, see Figure 3-15 for the example. 3D depth resolution also needs to be improved to benefit the recognition algorithms. All of these limitations suggest that the optimism sometimes expressed for 3D face recognition is still somewhat premature (Bowyer et al., 2004). Thus the appropriate issue may not be 3D versus 2D, but instead the best method to combine 3D and 2D.

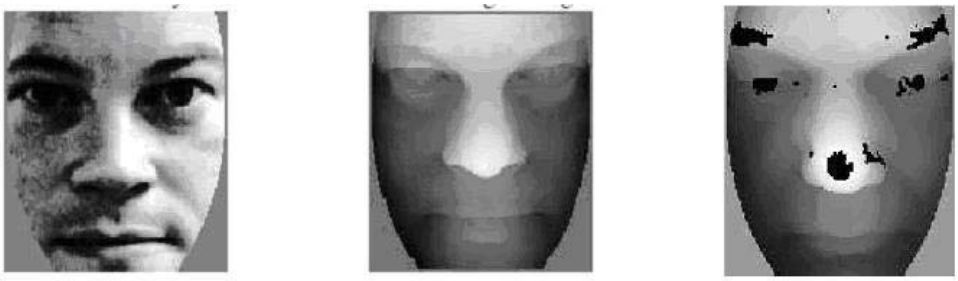


Figure 3-15 Example 2D intensity image, 3D range image and sample "Hole" in sensed 3D data (Bowyer et al., 2004)

3.4 Summary

A detailed survey of 2D face recognition algorithms and particularly, Gabor feature based methods has been given in this chapter. 3D face recognition approaches are also briefly described. The short survey on 3D approaches shows that 3D technologies are still at the initial stages due to a number of limitations. Texture, appearance, geometrical features etc. 2D information will continue to play important roles in face recognition. 2D methods can be basically classified into three categories: analytic, holistic and hybrid. While the analytic methods extract feature from prominent facial feature points, the holistic methods extract feature from the whole face pattern. Due to the robustness against complex distortions caused by various variations in illumination,

facial expressions and poses, Gabor wavelets seems to be promising basis to extract local features for face recognition, for several reasons:

- Biological motivation: the shapes of Gabor wavelets are similar to the receptive fields of simple cells in the primary visual cortex (Daugman, 1985),
- mathematical motivation: the Gabor wavelets are optimal for measuring local spatial frequencies (Kruger et al., 2002a; Kruger et al., 2002b), and
- empirical motivation: they have been found to yield significantly better performance than other methods in some performance tests (Zhang et al., 1997) (Kalocsai et al., 1998), FERET evaluation (Phillips et al., 2000) and FVC2004 competition (Messer et al., 2004).

Similar to general 2D face recognition algorithms, Gabor wavelet based approaches are also categorized as analytic and holistic methods. When elastic graph matching based analytic methods represent face images with different graph structures, the elastic matching process to locate graph nodes for a face image is however, very time consuming. To replace such a complex process, some researchers locate facial features by edges, colours etc. such that Gabor features can be extracted from those fiducial points for recognition. The location algorithm itself has to be robust against distortions caused by illumination, pose and expression. The success of Gabor feature based holistic methods relies on an augmented vector extracted from the whole face image, which is usually with huge dimension, e.g. 655,360 for image with size 128×128 when 40 wavelets are used. The feature thus requires high memory cost and could add high computation cost to the classifier as well. As a result, the research presented in this thesis will focus on application of Gabor wavelets for face recognition, and on developing methods to optimize the Gabor feature extraction process for performance improvement and computation/memory cost reduction.

Chapter 4 Gabor Features and Kernel Subspace Analysis for Face Identification

The detailed review on the background of Gabor wavelets has suggested the robustness of such mathematical tools for feature extraction. Once robust features are extracted, subspace analysis could be applied for further class separability enhancement and feature dimension reduction. Due to the adoption of kernel methods, non-linear kernel subspace analysis, e.g. Kernel Principal Component Analysis (KPCA) and Generalized Discriminant Analysis (GDA), might have substantial advantages over linear subspace techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). This chapter presents work that utilises Gabor features and kernel subspace analysis for face identification. A set of 40 Gabor wavelets is used to extract robust features, which are then subjected to KPCA or GDA to handle non-linear variations. Thereafter, different distance measures are evaluated and the nearest neighbour classifier is used for recognition.

4.1 The Methodology

4.1.1 System Architecture

Figure 4-1 shows a flow chart demonstrating the use of Gabor features and kernel subspace analysis for face recognition. Initially a set of Gabor wavelets are used to extract appropriate features, this process is detailed in the next section. The Gabor features extracted from a set of training images are then used to learn the kernel subspace, which is represented by the projection matrix $\mathbf{W}$. To identify a person, Gabor features of the face image are extracted, concatenated into a vector, projected to the learned kernel subspace and finally compared with the projections of training (gallery) images in the database. After comparison using a distance measure (such as Euclidean distance) the person is identified as the one whose image produces the smallest distance.

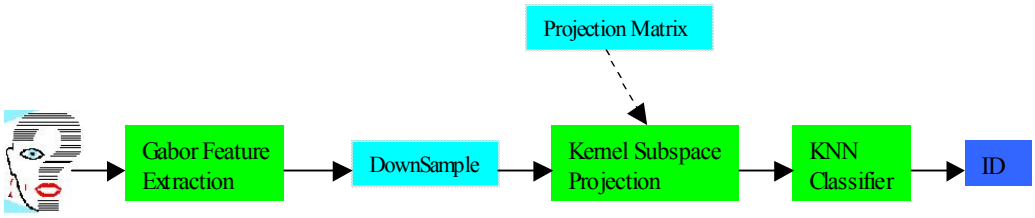


Figure 4-1 System architecture

4.1.2 Gabor Feature Extraction

As described in chapter 3, a Gabor wavelet is determined by the following parameters: the central frequency f, the orientation θ and the ratio between frequency and the sharpness of Gaussian axis γ,η. When the values of γ and η are normally fixed, a set of Gabor wavelets with different frequency and orientations should be designed to extract discriminant Gabor features. Most of the works in face recognition follow from the strategies proposed in (Lades et al., 1993; Wiskott et al., 1997), i.e.,

$\gamma = \eta = \sqrt{2}, f_u = {F_{\max}}/{\left(\sqrt{2}\right)^u}, \theta_v = {(v\pi)}/{8}$, where $u = 0,...,4, v = 0,...7$. Once the ratio is fixed, the size of the Gaussian envelope monotonically decreases with the value for the central frequency. The higher the central frequency of the Gabor sinusoidal carrier, the smaller the area the Gaussian envelop will cover in spatial domain. This is reasonable since the high frequency signal changes faster. According to the Nyquist sampling theory, a signal containing frequencies higher than half of the sampling frequency cannot be reconstructed completely. Therefore, the upper limit frequency for a 2D image is 0.5 cycles/pixel, while the lower limit is 0. However, for face images the actually useful band is much narrower, $F_{\max} = 0.25$ cycles/pixel has been proven to be a reasonable choice (Lades et al., 1993). A Gabor wavelet with parameters $(f_u, \theta_v, \gamma, \eta)$ can now be defined as:

$$\begin{aligned} &\varphi_{u,v}(x,y) = \frac{f_u^{\,2}}{\pi\gamma\eta}\exp\left(-\left(\left(\frac{f_u}{\gamma}\right)^2 x_r^{\,2} + \left(\frac{f_u}{\eta}\right)^2 y_r^{\,2}\right)\right)\exp\left(j2\pi f_u x_r\right) \\ &x_r = x\cos\theta_v + y\sin\theta_v \\ &y_r = -x\sin\theta_v + y\cos\theta_v \end{aligned}, \qquad (4.1)$$

Given a bank of 40 Gabor wavelets, $\{\varphi_{u,v}(x,y),\ u = 0,...,4, v = 0,...7\}$, image features at different locations, frequencies and orientations can be extracted by convolving the image $I(x,y)$ with the wavelets:

$$O^I{}_{u,v}(x,y) = \left|I(x,y) * \varphi_{u,v}(x,y)\right| \qquad (4.2)$$

Figure 4-2 shows the 40 Gabor wavelets and their representation in the frequency domain. As can be seen, the set of wavelets is tuned to a wide range of scales (frequencies) and orientations. The orientations of Gabor wavelet shown in the figure vary along the horizontal axis, while their scales vary in the vertical axis. The image in the 2nd row shows the spectrum of the 40 wavelets in frequency domain, with each blob representing the energy of a wavelet. To extract features at these different scale and

orientation levels the resultant Gabor feature set thus consists of convolution results of an input image $I(x,y)$ with all of the 40 Gabor wavelets:

$$S = \{O^I_{u,v}(x,y) : u \in \{0,...,4\}, v \in \{0,...,7\} \tag{4.3}$$

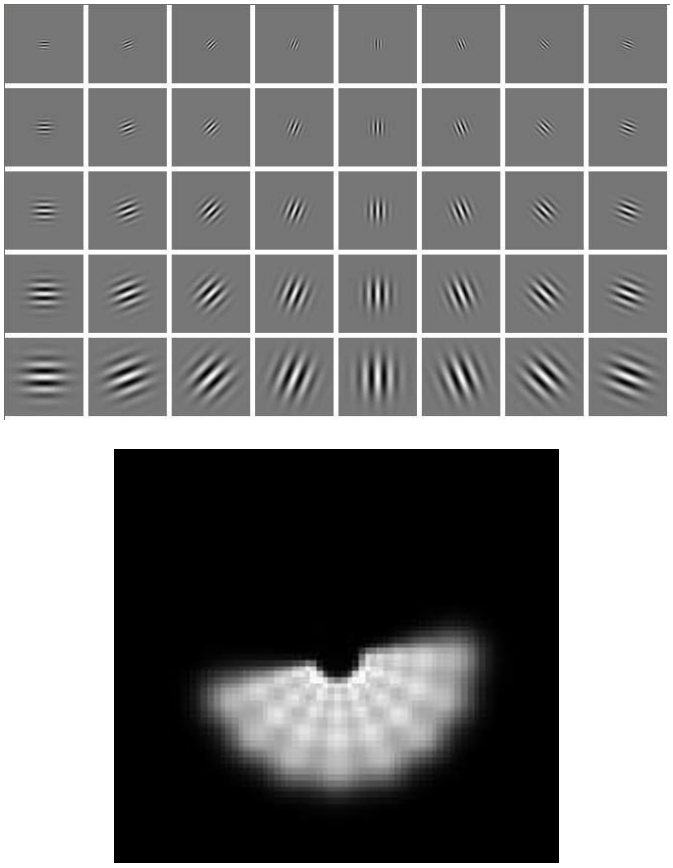


Figure 4-2 The 40 Gabor wavelets in the spatial and frequency domain

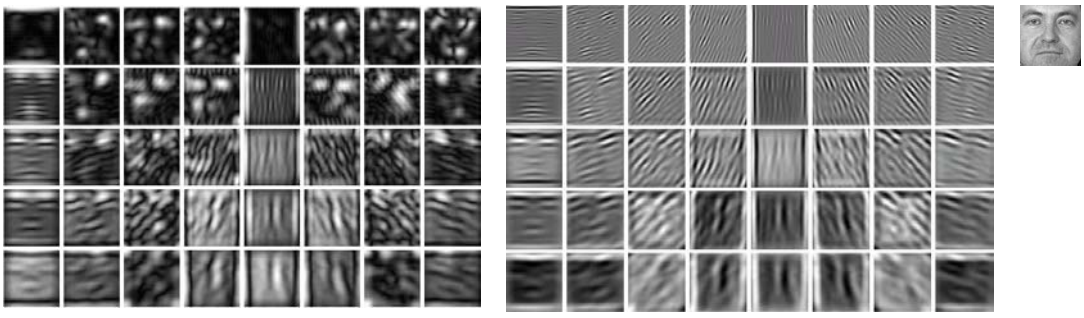


Figure 4-3 Convolution result - (magnitude and real part) of an image with 40 Gabor wavelets

Figure 4-3 shows the magnitude and real parts of Gabor representations of a face image at 5 scales and 8 orientations. A series of row vectors $\mathbf{O}_{u,v}^{I}$ could be converted out of $O_{u,v}^{I}(x,y)$ by concatenating its rows or columns, which are then concatenated together to generate a discriminative Gabor feature vector:

$$G(I) = \mathbf{O} = (\mathbf{O}_{0,0}^{I}\ \mathbf{O}_{0,1}^{I} \cdots \mathbf{O}_{4,7}^{I}) \tag{4.4}$$

As an example take an image with size 128×128, the convolution result will give 128×128×5×8=655,360 features.

4.1.3 DownSampling and Kernel Subspace Analysis

Due to the extremely high dimension of the extracted Gabor features, the computational cost associated with learning the subspace projection matrix is very high. Though the feature dimension does not affect the size of the kernel matrix it does increase the computational cost of the dot product of the two data samples. As suggested in (Liu et al., 2002), the Gaussian pyramid downsampling is used here for feature dimension reduction. The experiments of varying downsampling rate show that recognition rate drops drastically when the rate is larger than 64. However, the performance is actually very similar when the downsampling rate is less than 64. Considering both computation cost and system performance, the downsampling rate is set as 16 throughout this work. Take an image with size 128×128 for example, the dimension of Gabor features can now be reduced to 128×128×5×8/16=40,960.

Details of kernel subspace analysis have been discussed in chapter 3 where PCA and LDA are performed in the high dimensional feature space. By using the kernel technique the dot product of two data vectors in the mapped feature space can be easily computed from the kernel function. The KPCA and GDA subspace can thus be learned without knowledge of the mapping function. Due to its wide application in the radial

basis neural networks and Support Vector Machines the Gaussian kernel is used in this work:

$$k(x,y) = \exp\left(-\left\|x-y\right\|^2 \Big/ r\right) \tag{4.5}$$

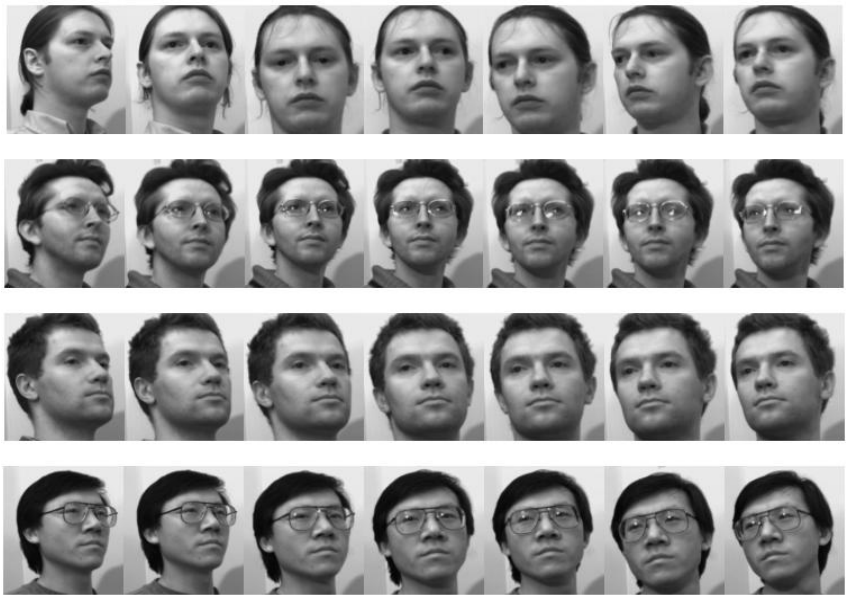


Figure 4-4 Sample images from the UMIST database

To give some initial ideas about the performance of kernel and linear subspace techniques, the ability of PCA, LDA, KPCA and GDA to separate data from different classes is considered first. In order to include non-linear variation within the sample set of face images the UMIST database (Granham & Allison, 1998) is used in this test. 128 face samples from 4 people (32 face images per person) are randomly selected. The database covers a range of poses from half profile to frontal views, see Figure 4-4 for the samples. Although the number of subjects is small in this example the variations of face images, even for the same person, are quite large. Due to the substantial pose variation the difference between face images of the same person might be larger than that due to the subject identity and thus the classification problem presented here is not a trivial one. The pixel values of the 128 training samples are directly used as features

and analyzed by PCA, LDA, KPCA and GDA respectively. The corresponding subspaces are constructed using the resultant eigenvectors. After that, the samples are projected onto the first two eigenvectors extracted by PCA, LDA, KPCA and GDA respectively. Figure 4-5 shows the distribution of the face samples in these subspaces after projection. In this example, the samples projected by LDA and GDA are well separated. The faces from the same person are projected to the same point by the GDA methods. This figure provides an example of better performance of LDA and GDA over PCA and KPCA. The discrimination ability of GDA is also proved in the experiments: GDA performs better than PCA, KPCA, and LDA when the FERET and ORL database are used for testing. The experimental results will be presented in detail in the following section.

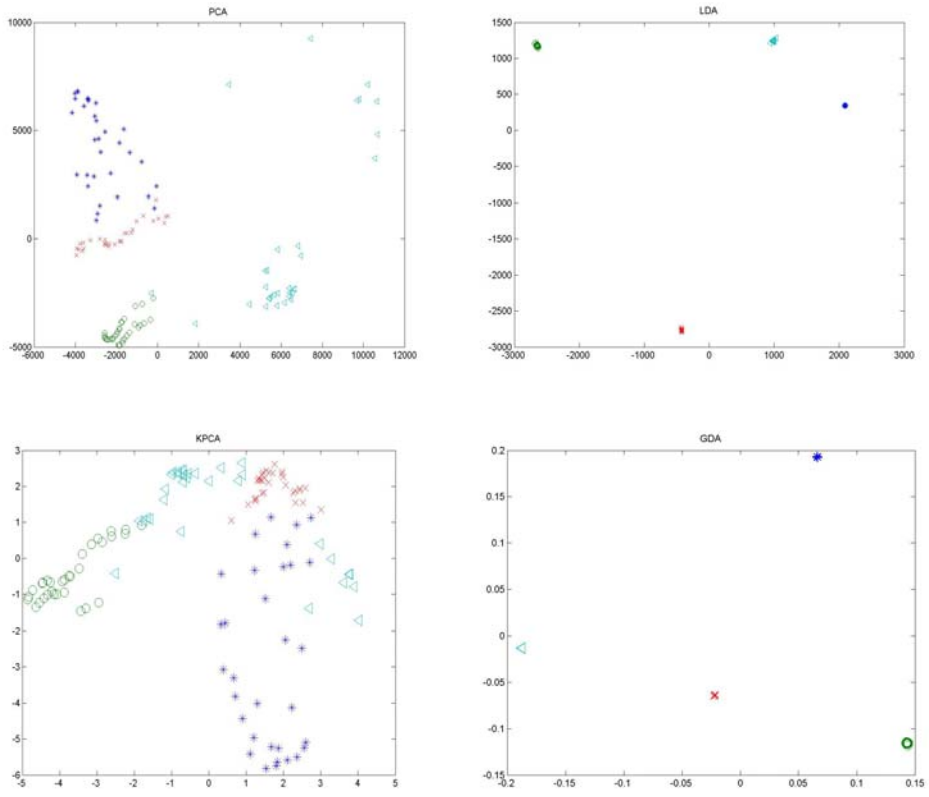


Figure 4-5 Distribution of face samples in PCA, LDA, KPCA and GDA subspaces

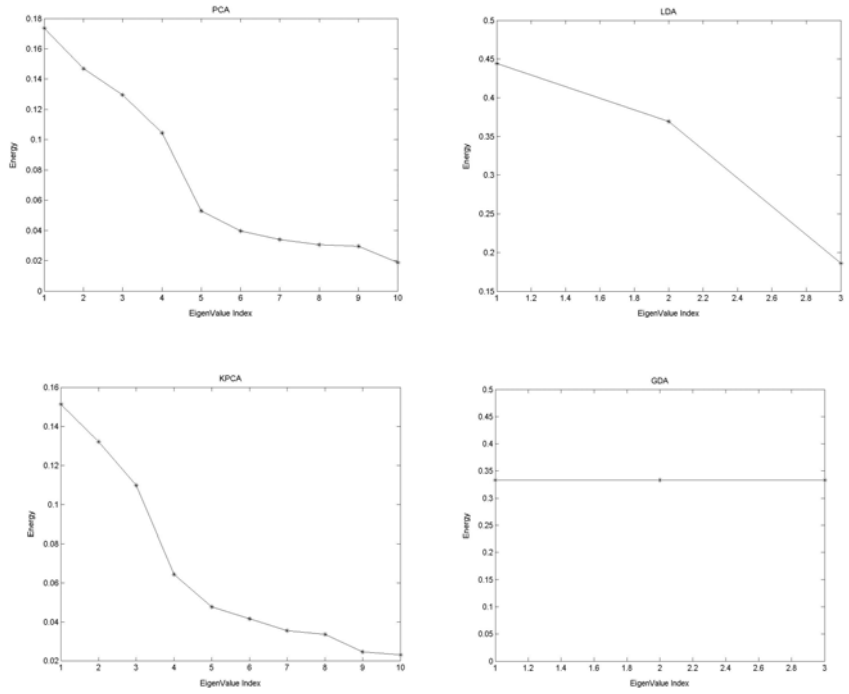


Figure 4-6 Energy of the eigenvalues in PCA, LDA, KPCA and GDA subspaces

Since the eigenvalues associated with learned projections (eigenvectors) might give important information to the discriminative ability of the subspace, we also show the energy of the eigenvalues for PCA, LDA, KPCA and GDA in Figure 4-6. Given a set of eigenvalues $\{\lambda_i, i = 1,2,\cdots,m\}$, the energy for λ_i is defined as $e_i = \lambda_i \Big/ \sum_i \lambda_i$. The maximum dimension of LDA and GDA is decided by $C-1$, where C is the number of individuals in the training set. As a result, while the energy for 10 eigenvalues are shown for PCA and KPCA, the energy for only 3 eigenvalues are shown for LDA and GDA. As shown in this figure, the variations of the eigenvalues of PCA and KPCA are quite similar in this example, which explains their similar classification performance. The eigenvalues shown for GDA are defined in the equation (A.4) in the appendix, which interestingly show that the first 3 eigenvalues are exactly the same. It seems that the 3 projections of GDA are equally important in this example.

4.1.4 Distance Measure and Classification

Given a set of training samples $\{x_i, i = 1,...,M\}$, a kernel function $k(x, y)$ and a subspace projection matrix $\mathbf{W}$ with dimension $M \times L, L << M$, a L dimensional feature $\boldsymbol{y}$ can be derived from the Gabor feature vector x extracted from a test face image by $\mathbf{y} = \mathbf{k}_x \mathbf{W}$, $\mathbf{k}_x = [k(x, x_1)\, k(x, x_2) \ldots k(x, x_M)]$. As described below, three different distance measures d_E, d_C, d_M are used in our experiments to calculate the distance between two sample projections y_1 and y_2:

Euclidean Distance (Eu):

$$d_E(y_1, y_2) = \sqrt{(y_1 - y_2)^T (y_1 - y_2)} \tag{4.6}$$

Mahalanobis Distance (Ma):

$$d_M(y_1, y_2) = (y_1 - y_2)^T \Sigma^{-1} (y_1 - y_2) \tag{4.7}$$

Normalized Correlation (Nc):

$$d_C(y_1, y_2) = \frac{y_1^T y_2}{\|y_1\| \|y_2\|} \tag{4.8}$$

where Σ is the covariance matrix calculated from the projected training samples, and $\|\cdot\|$ denotes the norm operator. The simple nearest neighbour classifier is used in our experiments for classification, i.e., the person is identified as the closest class to the input image:

$$i^* = \arg\min_{1 \le i \le M} d(y, y_i) \tag{4.9}$$

where M is the number of sample projections in the database.

4.2 Experimental Results

4.2.1 The Datasets

Now the performance of the Gabor feature and kernel subspace based methods are analyzed using two databases: the Face Recognition Technology (FERET) database

(Phillips et al., 2000) and the Olivetti Research Laboratory (ORL) database (The AT&T Lab Cambridge, 2002). The FERET database is associated with a testing procedure that is intended to evaluate face recognition systems. The facial images were collected in 15 sessions between August 1993 and July 1996. There are 14,126 images from 1,199 individuals included in the FERET database, which is divided into development and sequestered portions for evaluation. Due to the complexity of the Gabor feature based method, only a subset of the FERET database is used for testing in this chapter. However, with the improvements proposed in chapter 7, experimental results on the full FERET database according to the associated evaluation protocol will be given there. The ORL database contains face images taken between April 1992 and April 1994 at the University of Cambridge, UK. There are 400 images from 40 individuals.

The proposed method will first be tested using a subset of the FERET database, where variations in illumination and facial expression are available. Different distance measures for the Gabor + KPCA and Gabor + GDA methods will be evaluated and compared with the linear subspace techniques, i.e. Gabor + PCA and Gabor + LDA. The approach will also be compared with those using raw pixel values as features and state of the art algorithms in literature. Following the test on the FERET database, the proposed method will be further evaluated using the ORL database, where face images are captured with varied poses and scales. The performance will also be compared with that of state of the art techniques.

4.2.2 Performance Evaluation Using The FERET Database

600 frontal face images corresponding to 200 subjects are extracted from the FERET database for the experiments. All the subjects are in an upright, frontal position, with tolerance for some tilting and rotation of up to 10 degrees. The 600 face images were acquired under varying illumination conditions and facial expressions. Each subject has

three images of size 256×384 with 256 gray levels. The following procedures are applied to normalize the face images prior to the experiments:

- The centres of the eyes of each image are manually marked, each image is rotated and scaled to align the centres of the eyes,
- Each face image is cropped to the size of 128×128 to extract the facial region, and normalized to zero mean and unit variance.

To test the algorithms, two images of each subject are randomly chosen for training, while the remaining one is used for testing. Figure 4-7 shows sample images from the database. The first two rows are example training images while the third row shows example test images. One can see from the figure that all test images consist of variations in illumination and expression.

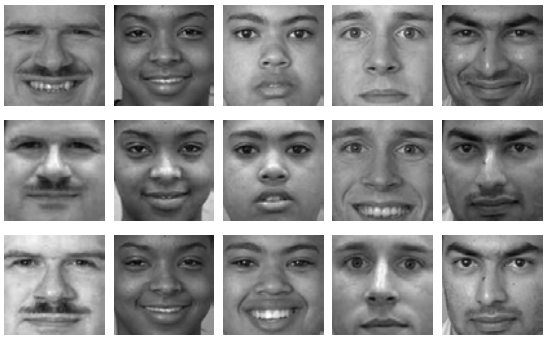


Figure 4-7 Example training images (top 2 rows) and test images (bottom row) of the FERET database

4.2.2.1 Comparison of Different Distance Measures

Kernel subspace analysis, i.e. KPCA and GDA, are performed on the Gabor feature vector extracted from the original face images for face identification. A Gaussian kernel is used for KPCA and GDA with $r = 8e4$, which is determined empirically for the best results, i.e. the value of r was chosen to maximize the recognition rate. We observe in our experiments that GDA is less sensitive to the value of r than KPCA. Three similarity measures Eu, Ma and Nr are tested and compared. As shown in Figure 4-8, normalized correlation achieved the best performance for GDA among the three

distance measures, while the difference between Euclidean distance and Mahalanobis distance is not large. However, Mahalanobis becomes the best distance measure for KPCA, which achieves significantly higher recognition rates than the other two measures (see Figure 4-9 for details). Similar results are also observed for the linear subspace projection methods, PCA and LDA. It seems that for expressive features derived in PCA and KPCA space, the Mahalanobis distance measure is more suitable than others; while for discriminating features extracted by LDA and GDA, the correlation distance measure seems to be the best choice.

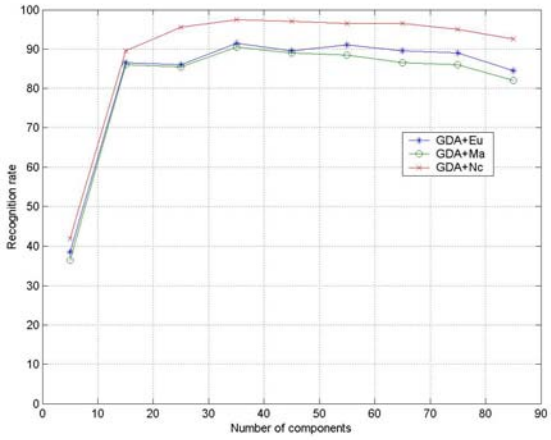


Figure 4-8 Performance of Gabor + GDA using different distance measures

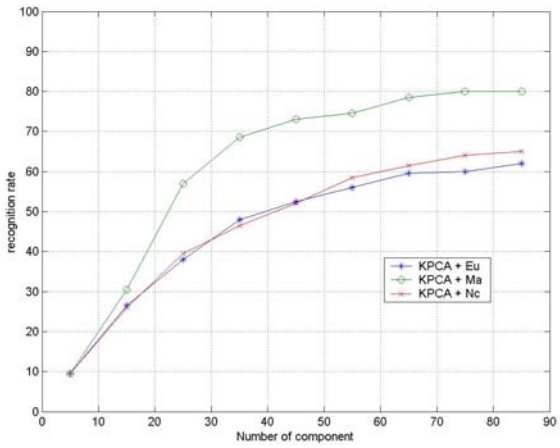


Figure 4-9 Performance of Gabor + KPCA with different distance measures

4.2.2.2 Comparison with Linear Subspace Methods

The comparative results of PCA, LDA, KPCA and GDA on the Gabor feature vector with respective optimized distance measures are shown in Figure 4-10. One can see from the figure that nonlinear subspace methods are basically performing better than their corresponding linear approaches, i.e., KPCA performs better than PCA and GDA performs better than LDA. GDA performs the best among these four algorithms. Following GDA, LDA performs better than KPCA and PCA. The results match well with the data separation test in section 4.1.3. A recognition rate as high as 97.5% is achieved for the novel Gabor + GDA approach when the number of components is set as 35. When the number of component became bigger than 90, we observed that the accuracy of PCA and KPCA converged around 80%, and there is no overlap between GDA and PCA, or KPCA.

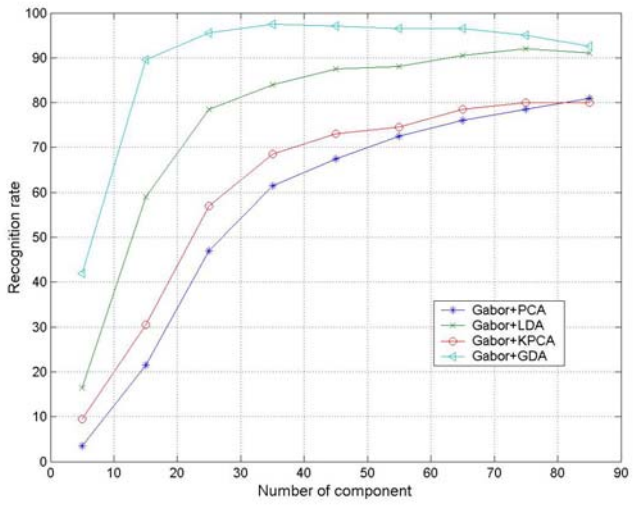


Figure 4-10 Experimental results of PCA, LDA, KPCA and GDA using Gabor features

4.2.2.3 Comparison with Raw Pixel Features

To emphasize the discriminating power of the extracted Gabor feature vector, the comparative performance of PCA, Gabor + PCA, GDA and Gabor + GDA are also shown in Figure 4-11. When the Gabor feature vector is not used, the pixel values of

face images are simply concatenated to a feature vector. For example, the length of a raw pixel feature vector will be 128×128=16,384 for an image with size 128×128. One can see that the adoption of the Gabor feature vector improves the performance of PCA and GDA by a large margin. The Gabor + PCA method achieves 20% higher accuracy than PCA, while 6% improvement is observed for GDA when Gabor wavelets are applied. The improvement for Gabor + LDA and Gabor + KPCA has also been observed in the experiments. Please note that the performance of GDA does not always improve with the increase of dimension. As the small (trailing) eigenvalues tend to capture noise, GDA achieves its maximum performance at dimension 35.

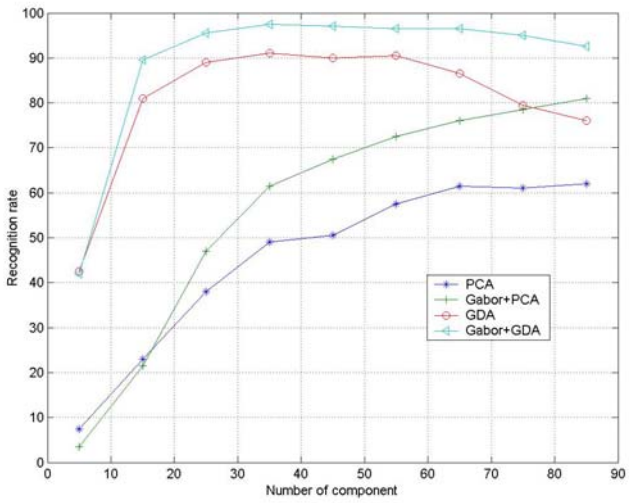


Figure 4-11 Performance improvement of PCA and GDA using Gabor features

4.2.2.4 Comparison with Other Methods

For further comparison of Gabor feature and GDA based methods with other approaches, the results on the same database for Radial Basis Function (RBF) neural network and HMM (Nefian et al., 1999; Bai et al., 2003a) based methods are shown in Table 4-1. Raw pixel features are used for RBF based methods, i.e. the normalized pixel values of the image are input directly to the network for personal identity determination. The two layers of the RBF network and HMMs are trained using the same training set,

with parameters optimized for best performance. The form of the neural network input layer is actually the Gaussian basis function, which is the same as the Gaussian kernel function. To make RBF the same structure with kernel subspace analysis, which takes inner product of the input data with all of the training samples, the network is designed with 400 nodes for the input layer and 200 nodes for the output layer. When DCT-HMM uses DCT (Discrete Cosine Transform) coefficients for observation vector extraction, DWT-HMM adopts DWT (Discrete Wavelet Transform) for more robust feature extraction. As shown on Table 4-1, Gabor + GDA performs significantly better than the other two methods.

	Recognition Rate
RBF Network	75%
DCT-HMM	32.5%
DWT-HMM	44.5%
GDA	90%
Gabor + GDA	97.5%

Table 4-1 Comparative results of Gabor + GDA with other methods on part of the FERET database

4.2.3 Performance Evaluation Using the ORL Database

The ORL database contains 400 images from 40 subjects. All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position, with tolerance for some tilting and rotation of up to 20 degrees. The variation in scale is up to about 10%. Figure 4-12 shows example training images and test images for 2 people. Each image is resized to 64×64 pixels and normalized to zero mean and unit variance. Both hair and forehead are included in the face images and the poses vary from left to right and up to down. To evaluate the algorithms, 5 images of each person are randomly chosen for training while the remaining 5 are used for testing.

PCA, LDA, KPCA and GDA are first performed on the original images for identification. As shown in last section, the Mahalanobis distance measure is used for

PCA and KPCA, while correlation distance measure is adopted for LDA and GDA. The results are tabulated in Table 4-2. As shown in this table, the performance of LDA deteriorates when the variation in pose increases the intra person variance significantly, thus it will be very difficult to find a projection space such that the within class variance is minimized. However, once the data is projected to the high dimensional feature space, GDA is still able to find the desired projection matrix. As a result, both PCA and KPCA achieve better performance than LDA, while GDA is still the best method for recognition.

(a)

(b)

(c)

(d)

Figure 4-12 Example training (a), (c) and test images (b), (d) in the ORL database

Method	Recognition Rate
PCA	92.0%
LDA	85.0%
KPCA	91.5%
GDA	96.5%

Table 4-2 Experimental results of PCA, LDA, KPCA and GDA on the ORL database

In the next series of experiments, PCA, LDA, KPCA and GDA are applied to the Gabor features extracted from the images and the results are shown in Table 4-3. The performance of LDA was greatly improved and it now achieves better performance than PCA and KPCA, which shows the robustness of Gabor features against the variation of pose. The novel Gabor + GDA methods achieve 100% accuracy when only 35 components are used, which is so far the best reported results in literature on the ORL database, see Table 4-4 for the results of other methods. Since the number of subjects in the ORL database is much smaller than FERET database, all algorithms achieve much better performance. The results are taken from the original papers directly, where the same testing strategy is used, i.e. half of the images are used for training and the remaining images are used for testing.

Method	Recognition Rate
Gabor + PCA	98.5%
Gabor + LDA	99.0%
Gabor + KPCA	98.5%
Gabor + GDA	100.0%

Table 4-3 Performance improvements using Gabor features on the ORL database

Method	Recognition Rate
RBF (Er et al., 2002)	98.08%
DCT-HMM (Bai et al., 2003a)	97.50%
DWT-HMM (Bai et al., 2003a)	98.50%

Table 4-4 Results of other methods on the ORL database

4.3 Conclusions

A Gabor feature and kernel subspace analysis based face identification method has been presented in this chapter. Gabor wavelets are used to extract features from the face images, which are then further analyzed by kernel subspace methods, such as, KPCA and GDA in order to achieve a highly discriminative feature for recognition. Two databases, FERET and ORL, have been used to test the proposed algorithms. While the face images extracted from the FERET database were acquired under variable

illumination and expressions, the samples in the ORL database represent variations in pose and scale. The results show that better performance can be achieved for kernel methods than their corresponding linear methods. By testing PCA, LDA, KPCA and GDA using the pixel features and the extracted Gabor feature vector respectively, the results show that the Gabor feature vector extracted from the filtered images yields a significantly more discriminative representation of the face than the original image. Comparison among different state of the art techniques show that the Gabor + GDA method achieves much more efficiency on both the FERET and ORL databases. As high as 97.5% and 100% accuracy have been observed on the two databases. By mapping the input features to a high dimensional nonlinear feature space, GDA can not only greatly reduce the feature dimension, but also increase the discrimination power of the extracted features. Encompassing different scale, locality and orientation information, the proposed Gabor + GDA method has bee proven to be very robust against variations of illumination, expression, pose and scale.

Chapter 5 Generalized Discriminant Analysis of Gabor Features for Face Verification

Whilst face identification aims to identify the personal ID of an input image, verification attempts to verify a claimed ID associated with a facial image. As a result, while an identification system needs to compare its input with each person in the database, verification systems attempt to match an input image with the claimed identity only. Based on the matching result, the system either accepts, or rejects the claimed ID. Applications of face verification can be found in passport control, E-business, personal authentication and in many additional areas.

Due to the successful application of Gabor features and GDA for face identification, this chapter presents a face verification system using the same technology. Robust Gabor features are first extracted from different face images, projected to the trained GDA subspace, and matched using the normalized correlation distance measure. The system will be fully tested using the BANCA database according to evaluation protocols of the recent Face Verification Competition 2004. As a result, the results are directly comparable with other participants.

5.1 Face Verification Competition 2004 and The BANCA Database

5.1.1 The Competition

With a large number of face recognition algorithms available in the literature, direct comparison between them is very difficult since tests are normally performed on different data sets. When images are captured with varying sensors, viewing conditions, illumination and backgrounds, it is unclear which method is the best. A standard test set with evaluation protocols could help alleviate this problem.

In Aug 2004, a face verification competition was organized by University of Surrey, UK. The contest was held in conjunction with the 17th International Conference on Pattern Recognition. 13 verification algorithms from 11 academic and commercial institutions around the world participated in the competition and the results are reported in (Messer et al., 2004). Different verification systems are first tested using face images normalized with manually located eye centres, and then assessed using their own automatic normalization methods. To make this work directly comparable with other participants, the verification methods presented in this chapter will be fully tested using exactly the same database and protocol as required by the contest.

5.1.2 The Database

Several data sets have been made available in literature over the past few years. While the FERET database (Phillips et al., 2000) defines a protocol for face identification evaluation, the XM2VTS database (Messer et al., 1999) can be used to test different face verification systems. The XM2VTS database, together with the Lausanne protocol, contains 295 subjects captured over 4 sessions. The data was recorded in a controlled environment, which makes it unrealistic compared to real world situations such as when one makes a transaction at home through a consumer web cam or through an ATM in a potentially very wide variety of surroundings. As a result, the BANCA database with

associated protocols (Baillere et al., 2003) has been proposed to make the evaluation as realistic as possible when real world factors are taken into consideration.

The BANCA database consists of images from 52 subjects captured in 12 sessions. 10 face images are captured for each person in each session. The 12 sessions are composed of 3 different scenarios: 1) Controlled scenario for sessions 1-4, 2) Degraded scenario for sessions 5-8, 3) Adverse scenario for sessions 9-12. A web cam was used in the degraded scenario and a high quality camera was used in the controlled and adverse scenarios. Images are captured with normal pose in the controlled and degraded scenarios, whilst a head down pose is required in the adverse scenario. Figure 5-1 shows some sample images captured in different scenarios from this database. All of the images are colour images with a size of 720×576. Images captured in different scenarios: controlled, degraded and adverse are shown on the first, second and third rows respectively.

Figure 5-1 Example Images in the BANCA Database

5.1.3 Test Protocols

Seven test protocols, which identify different training and test images, are defined in (Baillere et al., 2003) to evaluate verification algorithms. Of these protocols, protocol P is the most difficult and challenging one. The protocol specifies the partitioning of the database into two disjoint sets: a development set (26 subjects) and an evaluation set (26 subjects). For each set, 5 images from each person captured in the 1st session (Controlled scenario) are used as training images, while 2730 selected images captured in all three scenarios are used for testing. There is no overlap between the training images and test images. Of the test images, 1170 images are claimed with the true identity (client access) to test FR, while other images are claimed with a false identity (impostor access) to test FA. Each set thus consists of 130 training images, with the test data consisting of 1170 client accesses and 1560 impostor accesses (Baillere et al., 2003).

The performance of verification systems is normally assessed by the False Acceptance Rate (FAR) and the False Rejection Rate (FRR). These two measures are directly related, i.e. decreasing the number of false rejections will increase the false acceptance rate. The point at which FAR=FRR is known as Equal Error Rate (EER). The lower the value of the EER, the more reliable the system. EER can be used to measure the system performance where FAR and FRR are equally important, Weighted Error Rate (WER) is defined for weighted FAR and FRR as below:

$$\mathrm{WER} = \frac{\mathrm{FAR} + R \times \mathrm{FRR}}{1 + R} \tag{5.1}$$

where $R = C_{FA} / C_{FR}$ defines the cost ratio between FAR and FRR, 3 distinct cases can be defined to assess verification systems:

- $R = 0.1$, FA is an order of magnitude less costly than FR

- $R = 1$, FA and FR are equally costly
- $R = 10$, FA is an order of magnitude more harmful than FR

Obviously, EER is a special case where FA and FR are equally harmful. In order to meet the requirements of the contest, the results of Gabor + GDA for the 3 cases are reported in this chapter.

5.2 The Methodology

5.2.1 System Architecture

Figure 5-2 shows the flow chart of the described approach using Gabor features and GDA analysis for face verification. The GDA subspace, represented by the projection matrix $\mathbf{W}$, is first learned from the Gabor features extracted from a set of training images. The registered facial images of each person are then projected to the GDA subspace and projection coefficients are saved as templates in the database. To verify a claimed personal ID, the same process is applied to a given input image and the projection is compared with the stored projections of the person to be verified (the claimed ID). A decision could be made by a simple thresholding strategy, i.e., if the similarity is above or equal to the given threshold, the claim is accepted; otherwise it is rejected.

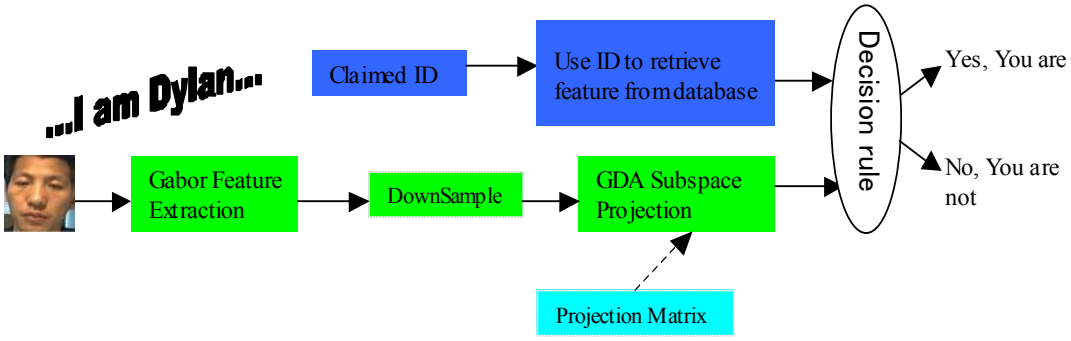


Figure 5-2 System architecture

5.2.2 Similarity Measure and Threshold Determination

Based on the evidence resulting from extensive experiments for the face identification approach, the robustness of the Gabor feature and GDA based methods were fully demonstrated in the previous chapter. The work also shows that the Mahalanobis distance measure should be used for expressive features such as PCA and KPCA, while the correlation distance measure is more appropriate for discriminative features derived by LDA and GDA. As a result, the work presented here follows this strategy. Since there might be a number of projections registered for a person i, the matching score, or confidence C_i of an input image projection y belongs to the subject is defined as below:

$$C_i = \frac{1}{N_i} \sum_{j=1}^{N_i} d(y, y_j) \qquad (5.2)$$

where N_i is the number of projections y_j registered for person i, and $d(y, y_j)$ are different distance measures such as Mahalanobis or correlation measures (see chapter 4 for details).

To make a decision on whether the claim is accepted, or rejected, a simple thresholding scheme can be used. While varying thresholds can be set for different people, a simpler approach is to use a global threshold for all of the subjects. A separate training set, or development set, can be used to determine the value of threshold(s). Thereafter, the performance of the system can be tested using a different test set. Whilst subject specific thresholds can achieve smaller error rates on the training or development sets, they might be easily over tuned to the training set and as a result, the simpler global threshold scheme is used throughout this work.

5.3 Experimental Results

5.3.1 The Dataset

To make the results of our method directly comparable with other methods in the competition, the BANCA database is used for testing in the experiments. Similar to the procedure used in chapter 4, all of the images used in the experiments are normalized semi-automatically. To achieve spatial normalization, face images are rotated, translated and scaled according to the position of the eyes. The images are cropped to a standard size of 48×48 and rotated so that the eyes are placed at fixed points. To reduce illumination variations, all of the images are initially histogram equalized and then shifted and scaled such that the mean values of all pixels equals zero, while the standard deviation equals one. While the results are reported on the manually normalized images in this chapter, results for the fully automatic verification system will be given in chapter 7. Figure 5-3 shows some normalized face images of three subjects acquired in different sessions: controlled, degraded and adverse scenarios are shown on the first, second and third rows respectively.

Figure 5-3 Normalized face images

5.3.2 Results on The Development Set

As defined in the protocol, a development set with 130 training images and 2730 test images from 26 subjects is first used to test the system. All the parameters of the system, e.g., subspace dimension, RBF kernel and decision threshold etc., are optimized to

maximize its performance on the development set. The results for Gabor + GDA are listed in Table 5-1, together with the baseline approach, Gabor + PCA. The reason behind the choice of PCA as a baseline is that LDA does not perform well when the training images are not representative, which is the case here since most of the test images are captured under distinct scenarios. Whilst PCA uses the Ma distance measure, the Nc distance measure is adopted for the GDA method. A RBF kernel with r=9e4 is found to achieve the best results. The ROC curves for the two methods using the development set are also shown in Figure 5-4. It can be seen from this figure that the Gabor + GDA method performs the best with a 5.96% EER (See Table 5-1). As described before, a global threshold is used for an acceptance or rejection decision.

Method	Kernel	Threshold	FAR	FRR	EER
Gabor + PCA	N/A	26.00	8.20	8.11	8.15
Gabor + GDA	RBF (r=9e4)	0.22	6.02	5.89	5.96

Table 5-1 Verification performance on the development set

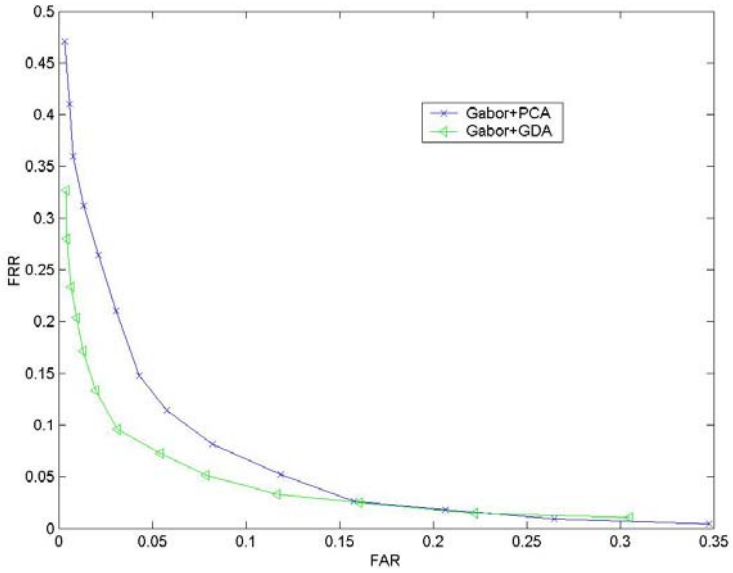


Figure 5-4 ROC curves on the development set

5.3.3 Results on The Evaluation Set

An independent evaluation set was designed in protocol P to test the generalization ability of the verification algorithms. The evaluation set consists of the same number of

subjects and images as that of the development set. However, the subjects of the evaluation set are distinct from those in the development set. With parameters adjusted and performance optimized using the development set, the generalization ability of algorithms can be further analyzed using the evaluation set. The EER of the Gabor + PCA and Gabor + GDA methods on the evaluation set are tabulated in Table 5-2. All of these results have been tuned to the development set in the first series of experiments, i.e., the decision threshold has been adopted during the development phase. Again, the Gabor + GDA method achieves a lower EER than Gabor + PCA. However, the advantage of GDA over PCA is not big in this test, which might be caused by the small size of the training set and the significant difference between the face images in the training set and the test set.

Method	Threshold	FAR	FRR	EER
Gabor + PCA	26.00	7.17	9.57	8.37
Gabor + GDA	0.22	7.75	7.43	7.69

Table 5-2 Verification performance on the evaluation set

5.3.4 Comparison with Other Methods

Once the performance of the Gabor + GDA approach has been analyzed using EER, it is now compared with all of the participants in FVC2004. Table 5-3 describes these results. Using the definition given in last section, the performance is now assessed using WER with 3 different values of *R*. Please note that the entry for "Univ Nottingham" is the other method developed by us, which uses Gabor wavelets for feature extraction, PCA for feature dimension reduction and Support Vector Machine (SVM) for classification. Since an executable exe file is required, we developed another method simply because of the insufficient of time available to convert the Gabor + GDA method into *C* implementation. Subject specific SVMs and thresholds are learned for each person. Once the parameters are optimized using the development set to achieve

the lowest possible WER, the same parameters can then be used when reporting the WER on the evaluation face image set. A number of different technologies have also been involved in this competition, e.g., PCA and LDA for feature extraction and dimension reduction, Hidden Markov Models (HMM) and Gaussian Mixture Models (GMM) for probability based classification and Nearest Neighbour (NN) and SVM for distance based classification. The IDIAP Fusion system is composed of three classification subsystems, i.e. DCT + HMM, DCT + GMM and LDA + Multi-layer Perceptron (MLP), the matching score of UCL-Fusion system is a weighted score of LDA + correlation distance and SVM. A more detailed description of the different approaches can be found in (Messer et al., 2004).

	R=0.1(WER)		***R=1(WER)***		***R=10(WER)***		**Avg**
	Dev	**Eval**	**Dev.**	**Eval.**	**Dev.**	**Eval.**	
IDIAP HMM	8.69	8.15	25.53	26.25	8.84	6.24	12.95
IDIAP Fusion	8.15	7.43	21.85	16.88	6.94	6.06	11.22
QUT	7.70	8.53	18.08	16.12	6.50	4.83	10.29
UPV	5.82	6.18	12.29	14.56	5.55	4.96	8.23
Univ Nottingham Gabor + PCA + SVM + subject specific thresholds	**1.55**	**1.77**	**6.67**	**7.12**	**1.32**	**1.58**	**3.33**
National Taiwan Univ	7.56	8.22	21.44	27.13	7.42	11.33	13.85
UniS	4.67	7.22	12.46	13.66	4.82	5.10	7.99
UCL-LDA	8.21	9.49	14.96	16.52	4.8	6.45	10.08
UCL-Fusion	6.05	6.01	12.61	13.84	4.72	4.10	7.80
NeuroInformatik	32.40	21.80	12.10	16.80	32.80	33.10	23.83
Tsinghua Univ	**1.13**	**0.73**	**2.61**	**1.85**	**1.17**	**0.84**	**1.39**
CMU	5.79	4.75	12.44	11.62	6.61	7.45	8.11
***Gabor + GDA +* global threshold**	***3.62***	***4.59***	***5.96***	***7.69***	***2.51***	***2.49***	***4.48***

Table 5-3 Verification results for partially automatic systems

The results for the proposed Gabor + GDA method have been appended to the bottom of the table. The comparison shows that the two methods developed by us are among the top three approaches. The performance of our methods have been shown to be significantly better than other participants except the Tsinghua University system,

which combines several classifiers for additional performance enhancement. Please note that the entry "NeuroInformatik" is based on the famous Elastic Bunch Graph Matching method (Wiskott et al., 1997), which extracts Gabor jets on manually defined feature points for recognition. Whilst their method achieves the top performance in the FERET evaluation (Phillips et al., 2000), it is has been shown not to perform to a high level of accuracy within the context of the FVC2004 competition. As specified in their description, their method may be more suitable for large and high quality images. Due to the adoption of subject specific thresholds, the method using Gabor + PCA + SVM achieves lower error rates than the Gabor + GDA approach presented in this chapter. However, methods using subject thresholds are more sensitive to the overfitting problem.

5.4 Conclusions

Following the successful application of Gabor + GDA methods for face identification, the same approach has also been used for solving the face verification problem. With very minor modifications, the system has proved to work well for verification applications. The system is fully tested using the BANCA database, which consists of images taken under uncontrolled environmental conditions. As a result, the test mirrors conditions found in real world application environments. By using the same database and protocol as the FVC2004, the results presented here are directly comparable with participants from all over the world. The comparison with other state of the art technologies shows that the work presented here is one of the most accurate, advanced and robust systems currently under development. With the exception to one institute, the method developed by us performs significantly better than other approaches. The results prove the robustness of the proposed Gabor feature and GDA subspace, thus the extracted features have been shown to be robust against variance of pose, illumination

and camera. Whilst the Tsinghua Univ combines several classifiers for performance enhancement, the second method developed by us uses subject specific thresholds. As a result, the performance of Gabor + GDA method could be further improved by fusing additional features and replacing global thresholds with subject specific ones. However, subject specific thresholds may cause the system to be over-tuned to the available data and thus a different data set may cause the system performance to drop dramatically, where as more generalised methods should naturally handle the change in data more appropriately.

Chapter 6 Optimising Gabor Features for Object Detection and Recognition

As shown in the previous chapters, the Gabor + GDA method has been successfully applied to both face identification and verification problems. The proposed approach was fully tested using the FERET and BANCA database and excellent performance has been observed. However, since a set of 40 Gabor wavelets is used to extract features, both computation and memory costs for this method are very high. The costs are mainly caused by the following processes: 1) the convolution operation of the image with 40 wavelets. Though FFT and IFFT can be used to speed up the process, the 40 convolution operations for a 128×128 image using a P4 1.8GHz PC still takes about 2 seconds; 2) the huge dimension of extracted features, i.e. 128×128×40 = 655,360 for a 128×128 image, brings a large memory and computation burden to the classification algorithm. A feature selection method, capable of reducing the number of convolutions and feature dimension, is required to solve such problems.

In this chapter, feature selection schemes such as AdaBoost algorithm will be applied for Gabor feature selection. The approach presented here aims to apply the optimal Gabor wavelets at the most appropriate locations for feature extraction. To reduce the redundancy among AdaBoost selected features, a novel boosting based feature selection algorithm --- MutualBoost is also proposed.

6.1 AdaBoost Feature Selection and Classifier Learning

The AdaBoost algorithm is based on the idea that a "strong classifier" can be created by linearly combining a number of "weak classifiers" (Freund et al., 1999). For image related problems, a weak classifier could be a very simple threshold function h_j consisting of only one simple feature $f_j(I)$ extracted from the image I:

$$h_j = \begin{cases} 1 & if \;\; p_j f_j(I) < p_j \lambda_j \\ -1 & \text{otherwise} \end{cases} \qquad (6.1)$$

where λ_j is a threshold and p_j is a parity to indicate the direction of the inequality. The feature could be the simple Haar-like features as described in (Lienhart et al., 2002), i.e., the linear combination of the sum of pixel values of neighbouring rectangles. Various features thus differ in any of the following rectangle parameters: location, width, height, and orientation $\alpha \in \{0, 45^0\}$. According to the structure of the neighbour rectangles, the features can be classified into 14 prototypes, i.e., four edge features, eight line features, and two centre-surround features. As shown in Figure 6-1, if one denotes the black and white rectangles as r_1, r_2 and the sum of pixels of a rectangle as $S(r)$, a Haar-like feature given any rectangle structure in an image I can be denoted as $f_j(I) = w_1 S(r_1) + w_2 S(r_2)$, where weights $w_1, w_2 \in R$.

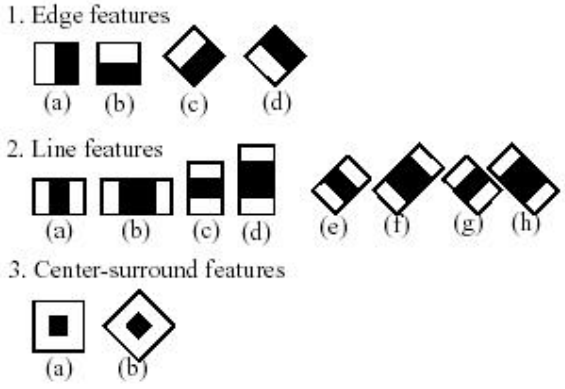


Figure 6-1 Prototypes of simple Haar-like features (Lienhart et al., 2002)

Details of the algorithm (also see chapter 2) are: T weak classifiers are selected to form the final strong classifier over a number of T rounds. In each of the iterations, the space of all possible classifiers is searched exhaustively to find the best weak classifier with the lowest weighted classification error. The error is then used to update the weights such that the wrongly classified samples get their weights increased. The resulting strong classifier is a weighted linear combination of all T selected weak classifiers. Since each weak classifier is using different features, the most important T features have also been selected. Note that AdaBoost algorithm is used here to address two class problems and weak classifiers with discrete output only. See AdaBoost.M1 and AdaBoost.MH (Freund et al., 1999) for solutions to the multi-class problem and RealBoost (Schapire & Singer, 1999) for boosting weak classifiers with real valued output.

6.2 The Proposed MutualBoost Algorithm

As described in previous section, the AdaBoost algorithm selects weak classifiers and adjusts sample weights based on the classification error. The motivation behind the weight adjustments is to change the distribution of samples such that the weak classifier selected at current round T is "uncorrelated" with the class label in the next round $T+1$ (Freund et al., 1999; Aslam, 2000). Intuitively, the learner is thus forced to learn something new in the next round $T+1$. However, a correlation between the class label and a certain weak classifier selected at round t, $0 < t < T$, might still exist. In this case, the weak classifier selected at round $T+1$ could be similar with the one selected at round t. As a result, many features selected by the AdaBoost algorithm might be similar (Li & Zhang, 2004).

The proposed boosting algorithm incorporates the idea of Mutual Information (MI) to eliminate those non-effective weak classifiers. Before a new weak classifier is added,

the MI between the new classifier and each of the selected ones is examined to make sure that the information carried by the new classifier has not been captured before. Given stage *T+1* where *T* weak classifiers $\{h_{v(1)}, h_{v(2)}, \cdots\ h_{v(T)}\}$ have been selected, the function to measure the max MI $R(h_j)$ between a candidate classifier h_j and the selected classifiers can be defined as follows:

$$R(h_j) = \max_t I(h_j, h_{v(t)}), t = 1,2,\cdots T \qquad (6.2)$$

Each weak classifier $h_j : R^N \to \{-1,1\}$ is now considered as a random variable (r.v.). the estimation of MI between two r.v., e.g. h_i and h_j requires information about the marginal distribution $p(h_i)$, $p(h_j)$ and the joint probability distribution $p(h_i, h_j)$, which could be approximated by histogram estimation. However, it is very difficult to determine the ideal number of histogram bins. Though a Gaussian distribution could be applied as well, many of the features, might not show Gaussianity. To reduce the complexity and computation cost of the feature selection process, we hereby focus on random variables with binary values only, i.e., $h_i \in \{-1,1\}, h_j \in \{-1,1\}$. For binary r.v., the probability could be estimated by simply counting the number of possible cases and dividing that number with the total number of training samples. For example, the possible cases will be $\{(-1,-1),(-1,1),(1,-1),(1,1)\}$ for the joint probability of two binary r.v. $p(h_i, h_j)$.

The value of $R(h_j)$ can be directly used to decide whether the new classifier is redundant or not. The value is compared with a pre-defined Threshold Mutual Information (TMI) value, if it is bigger than the TMI, we can deduce that the information carried by the classifier has already been captured. Besides MI, the classification error of the weak classifier is also taken in to consideration, i.e., only those classifiers with small errors

are selected. The classifiers (features) thus selected might be both accurate and non-redundant. Details of the algorithm are listed in Figure 6-2 as below.

Given M training samples $(x_i, y_i), i = 1,2,..,M$
Initialization: weights $w_1(i) = 1/M$
For *t=1, ..., T*
1) Train weak learners using distribution w_t
2) Given each candidate weak classifier h_j, calculate the classification error $\varepsilon_j = \sum_i w_t(i) |h_j(x_i) - y_i|$
   For (;;)
      Choose h_u with lowest error ε_u from the candidate classifiers
      Calculate the max MI $R(h_u)$ according to Eq. (6.2)
      If $R(h_u) < TMI$
         The classifier found, $h_t = h_u$, $\varepsilon_t = \varepsilon_u$
         go to 3)
      Else
         Remove h_u from the candidate list
      End If
   End Loop
3) Calculate $\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$
4) Update weights: $w_{t+1}(i) = \frac{w_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}$

Final strong classifier: $H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$

Figure 6-2 The proposed MutualBoost Algorithm

6.3 Application to Object Detection

Classification based object detection methods normally scan the image with a small window and make decision using a trained classifier as to whether the processing window is the object, or not. As described in section 1, AdaBoost algorithm has been successfully applied to select and learn Haar-like feature based classifier for object detection. In this system, each weak classifier is designed to make a prediction using single Haar feature extracted from image I, i.e. $f_j(I) = w_1 S(r_1) + w_2 S(r_2)$. In the context of Gabor feature selection, $f_j(I)$ is simply the convolution result of the input image with a

certain Gabor wavelet at location (x, y). Given an image with size $W \times H$ and a bank of $U \times V$ Gabor wavelets $\{\varphi_{u,v}(x, y), \quad u = 0,...,U-1, v = 0,...V-1\}$, a set of $N = W \times H \times U \times V$ Gabor features at different location, frequency and orientation can be extracted as below:

$$f_j(I) = \left(G(I)\right)_j, \quad j = 1,2,...,N \tag{6.3}$$

where $G(I)$ is the Gabor feature vector extracted from image I using the set of Gabor wavelets, i.e. $G(I) = (\mathbf{O}^I_{0,0}\ \mathbf{O}^I_{0,1} \cdots, \mathbf{O}^I_{u,v}, \cdots \mathbf{O}^I_{U-1,V-1})$. The row vector $\mathbf{O}^I_{u,v}$ is generated by concatenating the convolution result $\left|I * \varphi_{u,v}(x, y)\right|$ of an image I with a wavelet $\varphi_{u,v}(x, y)$, see chapter 4 for details. Each weak classifier is now trained to use a single feature from the complete Gabor feature set for classification. When these classifiers are combined, a much better performance can be achieved than that of single classifier.

Based on the importance of classification accuracy, essential Gabor features with appropriate frequencies and orientations are selected at different image locations and ranked by the AdaBoost algorithm. Once those discriminative Gabor features are selected, they can also be input to more complex classifiers, e.g. Support Vector Machine (SVM) for classification. The method will be applied to classify face/non-face and car/non-car images in the experiments and compared with Haar-like features based approaches. The Gabor feature based classifier can be further developed into a fast object detection system using a cascade structure as described in (Lienhart et al., 2002; Viola et al., 2001).

6.4 Application to Face Recognition

Since both algorithms of AdaBoost and the proposed MutualBoost are addressing two class problems only, the multi-class face recognition problem has to be reformulated to make the algorithms applicable. The Gabor feature difference space is adopted in this

work such that a set of training samples can be generated in the two class space. Once the set of samples and weak classifiers are available, Adaboost and MutualBoost can be applied directly for Gabor feature selection.

6.4.1.1 The Gabor Feature Difference Space

Since the feature selection presented here focuses on two class problems only, face recognition is formulated as a problem in the difference space (Phillips, 1999), which models dissimilarities between two facial images. Two classes, dissimilarities between faces of the same person (intra-personal space) and dissimilarities between faces of the different people (extra-personal space) are defined. The set CI (intra-personal difference) contains the within class difference, while the set CE (extra-personal difference) gives the dissimilarities among images of different individuals in the training set:

$$\begin{aligned} CI &= \left\{ \left| G(I_p) - G(I_q) \right|, I_p \sim I_q \right\} \\ CE &= \left\{ \left| G(I_p) - G(I_q) \right|, I_p \not\sim I_q \right\} \end{aligned} \tag{6.4}$$

where I_p and I_q are the facial images from people p and q respectively, and $G(\cdot)$ is the Gabor feature extraction operation as defined in last section. Each of the M samples in the difference space can now be described as $x_i = [g_1\, g_2 \cdots g_j \cdots g_N], i = 1,2,\cdots,M$, where N is the dimension of extracted Gabor features and $g_j = \left(\left| G(I_p) - G(I_q) \right| \right)_j$.

6.4.1.2 Training Samples Generation

For a training set with L facial images captured for each of the D persons, $D\binom{L}{2}$ samples could be generated for the intra-personal difference class while $\binom{DL}{2} - D\binom{L}{2}$ samples are available for extra-personal difference class. There are always many more extra-personal samples than intra-personal samples for face recognition problems. Take

a database with 400 images from 200 subjects for example, 200 intra-personal image pairs and $\binom{400}{2} - 200 = 79{,}800$ extra-personal image pairs are available. To achieve a balance between the numbers of training samples from the two classes, a random subset of the extra-personal samples could be produced. However, the generated subset should also be representative of the whole set. To achieve this trade off, the procedure as shown in Figure 6-3 is proposed to generate m extra-personal difference samples using $U \times V$ Gabor wavelets: instead of using only m pairs, the method randomly generates m samples from $m \times U \times V$ extra-personal image pairs. As a result, without increasing the number of extra-personal samples to bias the feature selection process, the training samples thus generated are more representative.

For $i = 1,2,\cdots m$
  For $u = 0,1,\cdots U-1$
    For $v = 0,1,\cdots V-1$
      Randomly generate an image pair (I_p, I_q) from different person
      Calculate the Gabor feature difference $\mathbf{Z}_{u,v}$ corresponding to filter $\varphi_{u,v}(x,y)$ using the image pair as below:

$$\mathbf{Z}_{u,v} = \left| \mathbf{O}_{u,v}^{I_p} - \mathbf{O}_{u,v}^{I_q} \right|$$

    End
  End
  Concatenate the $U \times V$ feature differences into an extra-personal sample, $x_i = [\mathbf{Z}_{0,0}\mathbf{Z}_{0,1}\cdots\mathbf{Z}_{u,v}\cdots\mathbf{Z}_{U-1,V-1}]$
End

Output the m extra-personal Gabor feature difference samples $\{(x_1, y_1), \cdots, (x_m, y_m)\}, y_1 = y_2 = \cdots = y_m = 1$.

Figure 6-3 Extra-personal difference samples generation

Including the $l = D\binom{L}{2}$ intra personal difference samples, the training sample generation process finally outputs a set of $M = m + l$ Gabor feature difference samples: $\{(x_1, y_1), \cdots, (x_M, y_M)\}$. Each sample $x_i = [g_1\, g_2 \cdots g_j \cdots g_N]$ in the difference space is

associated with a binary label: $y_i = -1$ for an intra-personal difference, while $y_i = 1$ for an extra-personal difference.

6.4.1.3 Weak Classifiers

Once a set of training samples with class labels (intra-person, or extra-person) $\{(x_1, y_1), \cdots, (x_M, y_M)\}$ is given, a large number of candidate weak classifiers h_j need to be designed for selection. Given a sample $x_i = [g_1\, g_2 \cdots g_j \cdots g_N]$ in the Gabor feature difference space, each weak classifier is now designed to be a simple threshold function using single feature, i.e., if the difference is less than a threshold, the prediction is set as -1, otherwise it is set as 1.

$$h_j = \begin{cases} -1, & \text{if} \quad \mathrm{g}_j < \lambda_j \\ 1, & \text{if} \quad \mathrm{g}_j \geq \lambda_j \end{cases} \tag{6.5}$$

Since we are only interested in the selection of features in this application, the threshold λ_j is simply determined by the centre of the intra-personal sample mean and extra-personal sample mean,

$$\lambda_j = \frac{1}{2}\left(\frac{1}{m}\sum_{p=1}^{m}\left((x_p)_j \middle| y_p = 1\right) + \frac{1}{l}\sum_{q=1}^{l}\left((x_q)_j \middle| y_p = -1\right)\right) \tag{6.6}$$

where m and l is the number of extra and intra personal difference samples, respectively. The set of candidate weak classifiers are now represented by N random variables with binary values, the MI between a candidate classifier and the selected classifiers can be easily calculated and the iterative process of MutualBoost as described in Figure 6-2 can be applied thereafter. On the other hand, the AdaBoost algorithm can be applied directly to the learned weak classifiers for selection.

The Gabor features thus selected by AdaBoost or MutualBoost are carrying important information about predicting whether the sample is an intra-personal difference, or an extra-personal difference. Based on the fact that face recognition is actually to find the

most similar match with the least difference, the selected features might be very important for recognition as well.

6.4.1.4 Kernel Enhancement

Once the most discriminative Gabor features are selected, they could be either used directly, or input to some classification system for face recognition. Different classification schemes could be used here, e.g., after Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA) is further applied for feature enhancement, the nearest neighbour (NN) classifier can be used for classification. In previous chapters, kernel subspace methods have been successfully applied to face identification and verification and the comparative identification results with linear subspace methods have clearly shown their advantage in handling nonlinear data. By mapping sample data to a higher dimensional feature space, effectively a nonlinear problem defined in the original image space is turned into a linear problem in the feature space (Scholkopf et al., 1999). Support Vector Machine (SVM) is another successful example of using kernel methods for classification. However, SVMs are basically designed for the two class problem.

Based on the successful application of Generalized Discrimniant Analysis (GDA) for face identification and verification in previous chapters, GDA is adopted here for further feature enhancement and KNN classification for recognition. The GDA subspace is first constructed from the selected Gabor features of training images and each image in the gallery set is then projected onto the subspace. To classify an input image, the selected Gabor features are extracted and then projected to the GDA subspace. The similarity between any two facial images can then be determined by the normalized correlation distance of the projected vectors. Details of applying GDA for face recognition can be found in chapter 4.

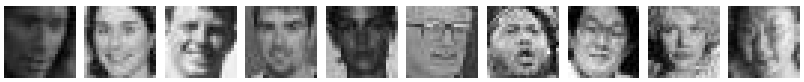
6.5 Experimental Results

6.5.1 Gabor Feature Based Classifier for Object Detection

The experiments presented here apply the AdaBoost algorithm to learning Gabor feature based classifier for object detection, which classify an image of standard size (e.g. 20×20 pixels) into either face (car) or non-face (non-car). As a two class problem, classification based methods (Rowley, Baluja, & Kanade, 1998; Osuna et al., 1997) have been one of the main approaches for object detection. Recent works (Lienhart et al., 2002; Viola et al., 2001) successfully built a face detection system with both high accuracy and fast speed. The system used the AdaBoost algorithm to select and learn Haar-like features based classifier for face detection. Following their framework, the experiments will perform two tasks: feature selection and classifier learning.

6.5.1.1 Data Sets

Two image sets, a face image set and car image set, are used to test the Gabor feature based object detection algorithm. The face image set is provided by Carbonetto (Carbonetto, 2001) and contains 4916 images with faces in them and 7872 images without faces in them. Figure 6-4 shows some example face and non-face images. All of the face images are of size 24×24, and are randomly split into a training set and test set containing 2458 positive samples (faces) and 3936 negative samples (non-faces) each. The second image set used in the experiments contains 550 images with at least one car in them and 500 images that do not contain a car (Agarwal, Awan, & Roth, 2002). The car image set is also randomly split into a training set and a test set. The training set contains 440 car images and 400 non-car images, whilst the remaining 110 car images and 100 non-car images are included in the test set. Figure 6-5 shows sample images from the car image set, which are of size 100×40.



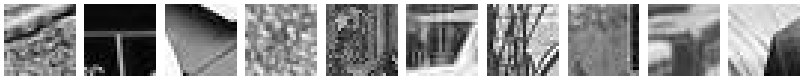


Figure 6-4 Images from face image set

Figure 6-5 Images from car image set

6.5.1.2 Selected Gabor Features

Given the set of the two classes of training samples with a class label, each sample could be represented with 24×24×40 = 23,040 Gabor features obtained by convolving 40 Gabor wavelets at each pixel location. Each Gabor feature obtained is thus associated with an image location and a Gabor wavelet. Once the most significant Gabor features to discriminate the two classes are selected by AdaBoost, their associated Gabor wavelets can be traced to gain information about the scale and orientation distribution of the wavelets. Figure 6-6 shows the distribution for the face image set. The scale with index $u, u = 1,2\cdots 5$ represent the wavelet with central frequency $f_u = {F_{\max}}/{\left(\sqrt{2}\right)^u}$ It is clear from the bar charts that the high frequency wavelets are chosen much more often than low frequency ones, and Gabor wavelets with orientation $\pi/2$ are preferred for this classification task. The orientation preference shows that horizontal features happen more frequently in face images, e.g. eyebrows, eyes, and mouth. The first eight Gabor wavelets selected by the AdaBoost algorithm for the car image set is also shown in Figure 6-7, which interestingly indicates that tyres are very important features for car detection.

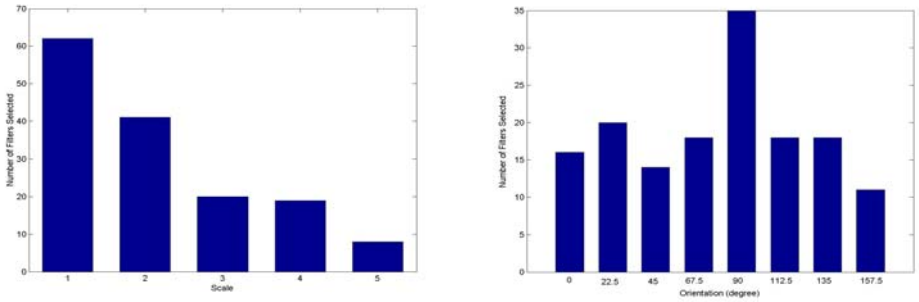


Figure 6-6 Scale and orientation distribution of filters selected for the face image set

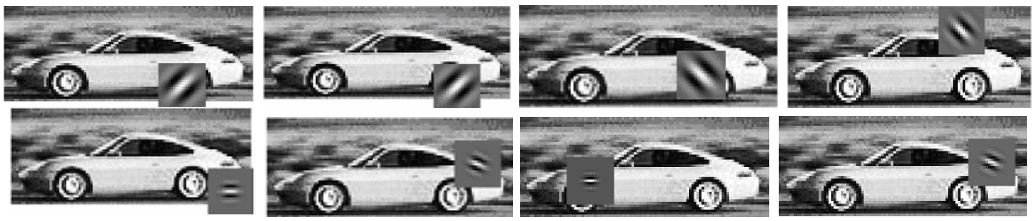


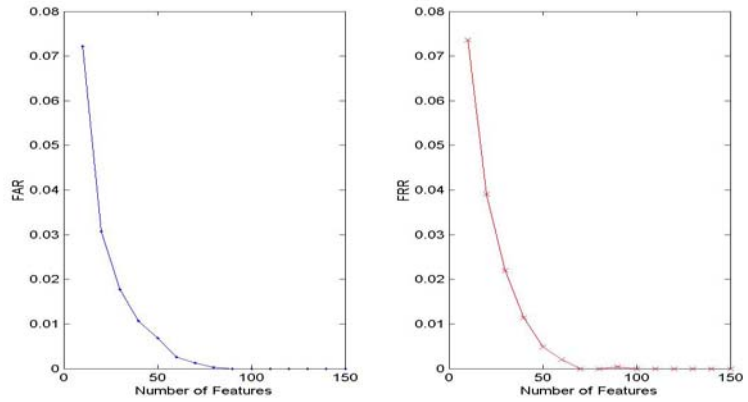
Figure 6-7 First eight selected Gabor wavelets for the car

6.5.1.3 Classification Performance Evaluation

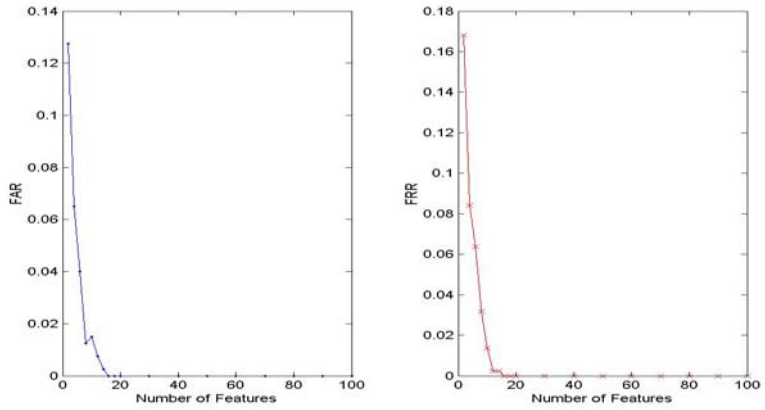
The AdaBoost algorithm not only selects the most discriminative Gabor features, but also learns a classifier using the selected features. The False Accept Rates (FAR) and False Reject Rates (FRR) for the AdaBoost trained classifier, GaborBoost, on the training image sets are shown in Figure 6-8. One can observe from the figure that 100 features are enough for GaborBoost to achieve zero FAR and FRR on the face image set, while only 20 features are required for the car image set. The results on the test face image set and the test car image set are shown in Figure 6-9. The best face/non-face classifier achieves 99.39% classification rate and 1.75% FRR with 150 selected Gabor features, while the best car/non-car classifier achieves 100% classification rate and 1.82% FRR with only 80 features.

To compare GaborBoost with other methods, the results of two other methods, named ExBoost and EABoost on the same face image set are also listed in Table 6-1. ExBoost uses the Haar feature set and AdaBoost algorithm to select features and learn classifiers, which is identical to the algorithm proposed in (Viola et al., 2001). They also proposed

to use a Genetic algorithm to reduce the search space during the boosting procedure, and named the algorithm as EABoost. As shown in the table, the GaborBoost algorithm outperforms ExBoost and EABoost in terms of both FAR and FRR, while using a fewer number of features. The results clearly show the advantages of Gabor features over Haar-like features in the context of object detection.



(a)

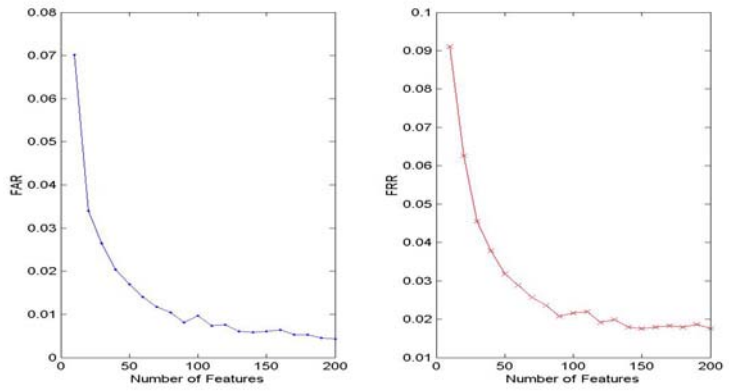


(b)

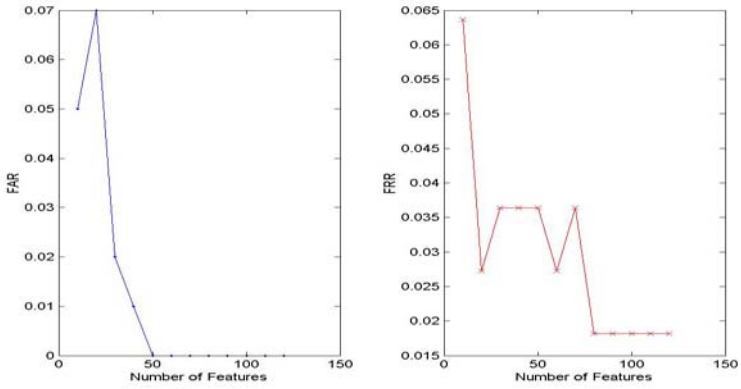
Figure 6-8 FAR and FRR on the training face image set (a) and the training car image set (b)

	Algorithm	Feature Numbers	FAR	FRR
Training Set	ExBoost	220	0	0
	EABoost	160	0	0
	GaborBoost	100	0	0
Test Set	ExBoost	227	3.9%	3.5%
	EABoost	163	3.1%	3.2%
	GaborBoost	150	0.61%	1.75%

Table 6-1 Comparative classification results on the face image set



(a)



(b)

Figure 6-9 FAR and FRR on the test face image set (a) and the test car image set (b)

6.5.1.4 SVM for Classification

In the following experiments, SVM is applied on the AdaBoost selected Gabor features for classification. The classifier, named as GaborBoostSVM, is trained using the Gabor features selected by the AdaBoost algorithm. Face images with the same partition of training set and test set are used for training and testing. 150 boosted Gabor features are extracted from each sample in the training set, which are then passed to SVM for training. The results are shown in Table 6-2 and compared with a SVM trained using the whole set of Gabor features with dimension 23,040 (GaborSVM), using the raw pixels (RawSVM) and GaborBoost as described above. For RawSVM, the pixel values

of each sample are concatenated to a feature vector to train a SVM. A Pentium 4 1.8 GHz PC and the SVM-Light package (Joachims, 2004) were used in our experiments.

	Gabor-Boost	SVM					
		GaborBoostSVM		**GaborSVM**		**RawSVM**	
		Linear	RBF	Linear	RBF	Linear	RBF
Feature Dimension	150	150	150	23,040	23,040	576	576
Number of SVs	N/A	233	271	503	N/A	1434	1386
SVM Training Time	N/A	38s.	75s.	10h.	>74h	180s	270s
FRR (%)	1.75	1.43	1.26	1.10	N/A	10.49	4.96
FAR (%)	0.61	0.36	0.30	0.18	N/A	3.78	0.97

Table 6-2 SVM classification results on the face image set

Compared with classifiers utilizing Gabor features, RawSVM achieves the highest FAR and FRR, which suggests that Gabor wavelets are a good choice for extracting features for classification. However, due to the huge dimension of Gabor features, we did not succeed in training GaborSVM using the RBF kernel - the program crashed after running for 74 hours, which may be caused by high memory usage and computation cost. It also takes about 10 hours to train the GaborSVM with a linear kernel. The training time appears to increase exponentially with the number of training samples. In addition, the computational cost of convolving an image with 40 Gabor wavelets is very high, which makes GaborSVM unsuitable for real time applications. Since the SVM is specially suited for binary classification, GaborBoostSVM achieves lower FAR and FRR than GaborBoost. Both methods use the same 150 Gabor features selected by the AdaBoost algorithm. The training of GaborBoostSVM with a RBF kernel takes less than 2 minutes. Only 150 convolution operations using one variable wavelet is necessary to extract the selected Gabor features, which makes GaborBoostSVM highly effective in terms of memory and computational efficiency.

6.5.2 Selecting Gabor Features for Face Recognition

Based on the discriminative power of Gabor features for pattern classification, the experiments presented in this section aim to learn the most significant Gabor features for face recognition. By reducing the feature dimension, not only is memory and computation cost greatly reduced, the system may also be more robust against the inference of noise. As a standard test bed, the FERET database (Phillips et al., 2000) is used here to evaluate the performance of selected Gabor features for face recognition. The same subset (600 frontal face images corresponding to 200 subjects) used in chapter 4 is first used here to compare the performance of different feature selection schemes, i.e. AdaBoost and MutualBoost. All of the images are normalized in both size (64×64) and orientation according to the eye coordinates. Both the difference between selected features and recognition performance will be analyzed. The recognition performance using the selected Gabor features will also be compared with the method shown in chapter 4, where the whole set of Gabor features before selection is used for identification. Once an improved feature selection approach for face recognition is identified, it will be applied to the whole FERET database according to the specified evaluation protocol for identification. Finally the performance will be compared with other state of the art algorithms.

6.5.2.1 Selected Gabor Features

The randomly selected 400 face images (2 images for each subject) are first used to learn the most important Gabor features for intra-personal and extra-personal face space discrimination. As a result, 200 intra-personal face difference samples and 1,600 extra-personal face difference samples using the method as described in Figure 6-3 are randomly generated for feature selection. Figure 6-10 and Figure 6-11 show the first six locations of the first 200 Gabor features selected by AdaBoost (AdaGabor) and

MutualBoost (MutualGabor) respectively, both are overlapped with a typical face image in the database. It is interesting to see that most of the selected Gabor features are located around the prominent facial features such as eye brows, eyes, noses and chin, which indicates that these regions are more robust against the variance of expression and illumination encountered within the database subset. This result is agreeable with the fact that the eye and eyebrow regions remain relatively stable when a person's expression changes. Though the first six Gabor wavelets selected by the AdaBoost and MutualBoost algorithms are similar, the locations of the 200 features show the existence of redundancy among AdaBoost selected features, i.e. many of the features are very near, or similar, to each other. The features selected by MutualBoost are more widely spread and thus exhibit a lower degree of correlation.

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

Figure 6-10 First six Gabor features (a)-(f); and the 200 feature points (g) selected by AdaBoost

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

Figure 6-11 First six Gabor features (a)-(f); and the 200 feature points (g) selected by MutualBoost

Figure 6-12 shows the distribution of MutualBoost selected wavelets in different scales and orientations. As shown in this figure, wavelets centred within low frequency bands are selected much more frequently than those in high frequency bands. On the other hand, the majority of the discriminative Gabor features have an orientation around $\pi/4$, $3\pi/8$, $\pi/2$ and $5\pi/8$. It is interesting to compare the two distributions of Gabor wavelets

selected for face detection and recognition: while the dominant orientations of the selected wavelets are similar for both applications, the dominant frequency bands are different – one prefers high frequency information and the other favours lower frequencies. This suggests that high frequency features are more important to discriminate objects with backgrounds. Since the differences between face images are used to select Gabor features for recognition, low frequency features seem to be more robust against the distortions caused by expression and illumination variations.

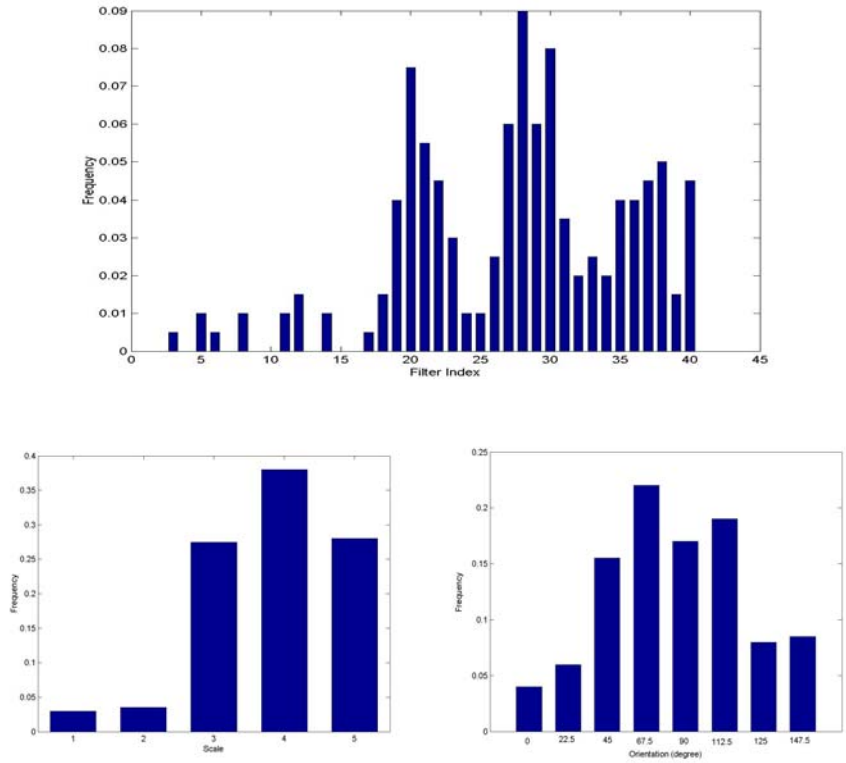
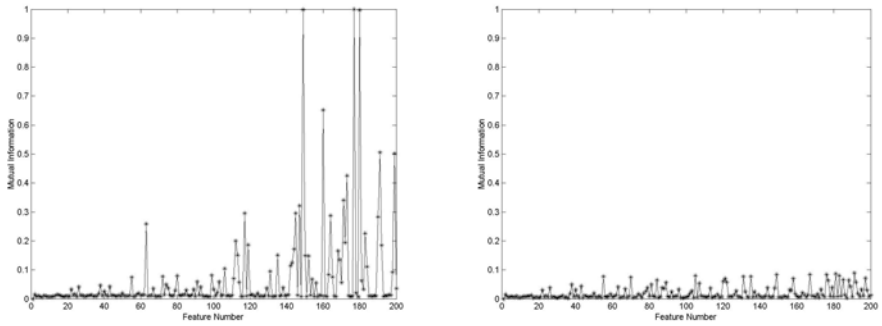


Figure 6-12 Distribution of MutualGabor features in scale and orientation

To show the existence of redundancy among AdaBoost selected features (weak classifiers), the max MI $R(h_j)$ for each selected feature is shown in Figure 6-13a. It can be observed from the figure that some of the features are highly redundant, e.g. the MI of features with numbers 149, 177 and 180 is greater than 0.99. The redundancy among selected features increases with the number of features, it is this undesired redundancy

that we aim to eliminate or reduce. The MI data for features selected with MutualBoost has also been shown in Figure 6-13b (with TMI=0.1). Due to the introduction of TMI, all the selected features now show MI values of less than 0.1 and thus one can conclude that the features are informative and non-redundant.



(a) (b)

Figure 6-13 MI of features selected by AdaBoost (a); MutualBoost (b)

6.5.2.2 Algorithm Complexity

Due to the introduction of mutual information, MutualBoost requires longer training time than that required by AdaBoost. However, the only computation cost added to AdaBoost is the loop to calculate MI values for redundancy checking, see Figure 6-2 for details. Table 6-3 shows the Average Number of Loops (ANL) required in each iteration and the corresponding TMI. The table shows that the computation burden added by the introduction of MI is actually very low (ANL is normally less than 10). As a result, the training time required by the proposed algorithm in our experiments is only about 0.1 times greater than that of AdaBoost.

TMI	0.08	0.09	0.10	0.11	0.12
ANL	8.42	8.07	7.25	5.43	3.25

Table 6-3 ANL for different TMI

As seen from the table, the higher the value of TMI, the less ANL required, i.e. the faster training speed. Actually AdaBoost can be seen as a special case of MutualBoost

when the value of TMI is set as 1. In this case, the features, or weak classifiers selected by the proposed algorithm will be exactly the same as those chosen by AdaBoost.

6.5.2.3 Recognition Performance on Subset of the FERET Database

Once different sets of Gabor features are selected, they can be used either directly, or subjected to further analysis for recognition. To compare the performance of different feature selection schemes, Both AdaGabor and MutualGabor are first applied directly for face recognition, with the resulting performance shown in Figure 6-14. The features were tested using 200 images (one for each subject), which are different from the training images in both illumination and expression. The normalized correlation distance measure and the nearest neighbour classifier are adopted. The performance shown in Figure 6-14 proves the advantage of MutualGabor over AdaGabor, i.e. the accuracy of MutualGabor is equivalent with, or higher than AdaGabor with any number of features. Since the MI values for all of the first 60 features are quite small, MutualBoost starts by picking up much the same features as AdaBoost. However, once the number of features increases, AdaBoost starts to pick redundant features. The improved recognition rate accuracy over AdaBoost caused by the use of features selected using MutualBoost shows the usefulness of the techniques in eliminating redundancy. The performance drop using 160 MutualGabor features could be caused by the variance between test images and training images - some features significant to discriminate training images might not be the appropriate ones for test images. A more representative training set might alleviate this problem. As shown in the figure, MutualGabor achieved as high as 94% recognition rate with 200 features.

In the next series of experiments, GDA will be performed on the selected Gabor features (MutualGabor + GDA) for further enhancement. To show the robustness and efficiency of the proposed methods, the performance of GDA on the whole Gabor

feature set (Gabor + GDA) is also included for comparison purposes. Downsampling is adopted to reduce feature dimension to a certain level, see chapter 4 for details. The normalized correlation distance measure and the nearest neighbour classifier were used. As described in chapter 4, the maximum dimension of the GDA subspace is determined by the number of classes and the number of non-zero eigenvalues of the kernel matrix. The maximum dimensions for Gabor + GDA and MutualGabor + GDA in this test are 110 and 199 respectively. As shown in Figure 6-15, MutualGabor + GDA achieves as high as 99.5% accuracy. Since all of the face images in this experiment are normalized to a reduced size (64×64) to speed up the feature selection process, the performance (97%) of Gabor + GDA is a little bit lower than that reported in chapter 4 (97.5%), which was tested on images of size 128×128. The performance improvement of MutualGabor + GDA shows that some important Gabor features may have been lost during the dowsampling process for Gabor + GDA. Additionally some of the remaining features are redundant.

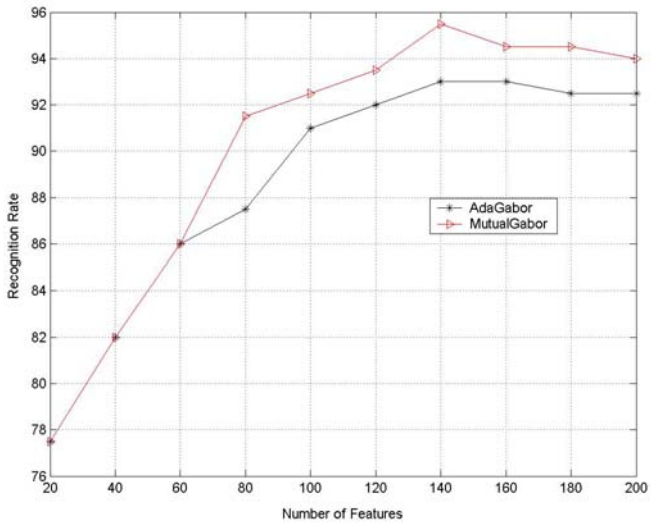


Figure 6-14 Recognition performance of AdaGabor and MutualGabor

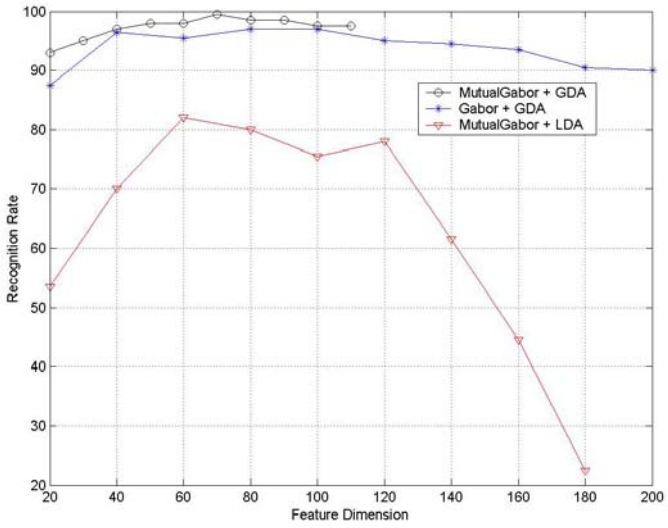


Figure 6-15 Recognition performance of enhanced MutualGabor

The computation and memory costs of Gabor + GDA and MutualGabor + GDA are also listed in Table 6-4. This shows that MutualGabor + GDA incurs significantly less computation and memory costs than Gabor + GDA, e.g., the number of convolutions to extract Gabor features is reduced from 16,3840 to 200. Although the Fast Fourier Transform (FFT) could be used here to circumvent the convolution process, the feature extraction process still takes about 1.5 seconds for images with size 64×64 in our *C* implementation whilst the 200 convolutions take less than 4ms. For Gabor + GDA with a down-sampling rate of 16, the feature dimension is reduced to 10,240, which is still 50 times the dimension of MutualGabor + GDA. As a result, MutualGabor + GDA is much faster in training and testing. While it takes Gabor + GDA 275 seconds to construct the GDA subspace using the 400 training images, it takes MutualGabor + GDA only about 6 seconds. MutualGabor + GDA also achieves substantial improvements to recognition efficiency - only 4 seconds are required to recognize the 200 test images. The computation time is recorded in Matlab 6.1, with a P4-1.8GHz PC.

With non-redundant and informative Gabor features, MutualGabor + GDA achieves better accuracy with significantly less computation than other methods described here.

	Number of Convolutions to Extract Gabor Features	Dimension of Gabor Features before GDA	Training Time	Test Time
Gabor-GDA	16,3840	10,240	275 sec.	263 sec.
MutualGabor -GDA	200	200	6 sec.	4 sec.

Table 6-4 Comparative computation and memory cost of Gabor + GDA and MutualGabor + GDA

Having shown in chapter 4 that GDA achieves significantly better performance on the whole Gabor feature set (Gabor + GDA) than LDA (Gabor + LDA), the performance of LDA on the selected informative Gabor features (MutualGabor + LDA) is also included in Figure 6-15 for comparison. As shown in the Figure, the performance of MutualGabor + LDA is substantially worse than that of Gabor + GDA and MutualGabor + GDA. Only 82% accuracy is achieved when the dimension of LDA subspace is set as 60, which is even worse than that of MutualGabor --- application of LDA surprisingly deteriorates the performance of MutualGabor. The result suggests that when the input features are discriminative enough, LDA analysis may not necessarily lead to a more discriminative space. The results also show that the feature enhancement ability of GDA is better than LDA.

6.5.2.4 Recognition Performance on the Full Set of FERET Database

After showing the comparative results with a state of the art Gabor feature based algorithm, the MutualGabor + GDA algorithm is now tested on the whole FERET database. According to the evaluation protocol, a gallery of 1196 frontal face images and 4 different probe sets are used for testing. The numbers of images in different probe sets are listed at Table 6-5, with example images shown in Figure 6-16. Fb and Fc probe sets are used for assessment of the effect of facial expression and illumination changes

respectively, and there is only a few seconds between the capture of the gallery-probe pairs. Dup I and Dup II consist of images taken on different days from their corresponding gallery images, and particularly, there is at least one year between the acquisition of the probe image in Dup II and the corresponding gallery image. A training set consisting of 736 images, is used to select the most informative Gabor features and construct the GDA subspace. Note that the same set was released to researchers to develop their algorithms during FERET evaluation. As a result, 592 intra-personal and 2000 extra-personal samples are produced to select 300 Gabor features using the sample generation algorithm and information theory. During the development phase, the training set is randomly divided into a gallery set with 372 images and a test set with 364 images to decide the dimension for optimal GDA performance. The same parameters developed are used throughout the testing process.

Probe Set	Gallery	Probe set size	Gallery size	Variations
Fb	Fa	1195	1196	Expression
Fc	Fa	194	1196	Illumination and Camera
Dup I	Fa	722	1196	Time gap < 1 week
Dup II	Fa	234	1196	Time gap > 1 year

Table 6-5 List of different prob sets

Figure 6-16 Examples of different probe images

Performance results of the proposed algorithm are shown in Table 6-6, together with that of the other main approaches participating in the FERET evaluation (Phillips et al., 2000), as well as an approach to extract Gabor features from variable feature points for recognition (Kepenekci, Tek, & Akar, 2002). The results show that MutualGabor +

GDA achieves the best result on all of the test sets. This can be attributed to the robustness of the selected Gabor features against variation in expression and capture time. Particularly, the performance of the proposed method is significantly better than all other methods on the Dup II set. Following the proposed method, the Elastic Bunch Graph Matching (EBGM) method, which is based on elastic graph matching, ranked as the second performer. However, the method requires intensive computation complexity for both Gabor feature extraction and graph matching. It was reported in (Wiskott et al., 1997) that the elastic graph matching process took 30 seconds on a SPARCstation 10-512. Compared with the EBGM approach MutualGabor + GDA is far superior in terms of both accuracy and computational efficiency.

Method	Fb	Fc	Dup I	Dup II
PCA	83.4%	18.2%	40.8%	17.0%
PCA + Bayesian	94.8%	32.0%	57.6%	35.0%
LDA	96.1%	58.8%	47.2%	20.9%
Elastic Graph Matching	95.0%	82.0%	59.1%	52.1%
Variable Gabor Features (Kepenekci et al., 2002)	96.3%	69.6%	58.3%	47.4%
MutualGabor + GDA	**96.7%**	**85.6%**	**59.3%**	**62.4%**

Table 6-6 FERET evaluation results for various face recognition algorithms

6.6 Conclusions

Two different algorithms: AdaBoost and the proposed MutualBoost have been successfully applied for Gabor feature selection in this chapter. The AdaBoost algorithm is used to learn Gabor feature based classifiers for object detection. While accuracy advantages of Gabor features over Haar-like features are observed using the AdaBoost learned classifier, further improvements have been achieved when SVM is adopted for classification. Due to the greatly reduced feature dimension, the SVM classifier using selected Gabor features achieves a substantial speed advantage over systems using the whole Gabor feature set. Based on its high accuracy, the module can

be further developed to a classification based object detection system. A cascade structure could be used to achieve a trade off between accuracy and efficiency.

The two feature selection schemes described have also been successfully applied to select Gabor features for face recognition. To simplify the computation cost and algorithm complexity, the intra-personal and extra-personal difference spaces are used. Compared with AdaBoost, experimental results show that features selected when mutual information is considered achieve higher recognition accuracy. The MutualBoost selected Gabor features are further enhanced in the non-linear kernel space using Generalized Dsicriminant Analysis and fully tested with extensive databases. Compared with one of the top methods in FVC 2004, the method shows advantages in both accuracy and efficiency. The results on the full FERET database following the evaluation protocol also show that the algorithm performs better than the previous top method, the elastic graph matching algorithm. However, the algorithm shows advantages in computation cost and efficiency since no graph matching process is needed. In addition, the method achieves significantly better performance on the most difficult test set, Dup II.

Whilst the mutual information based feature selection process in this chapter addresses the r.v. with binary values only, it could certainly be extended to the case of continuous variables. A Gaussian mixture model may be needed to represent the distribution when the r.v.s do not show Gaussianity. The distribution could also be discretized using histogram estimation, if the number of bins could be determined. When a r.v. with multiple values is used, the feature selection process will incur a much higher computation cost and complexity.

The value of TMI for MutualBoost needs to be selected appropriately to make sure that selected features are both non-redundant and useful for classification. A cross-

validation set could be used to determine the TMI for common classification problems. As shown in Figure 6-13, since the redundancy increases with the number of selected features, an adaptive TMI, which increases with the number of features, might be more suitable.

Chapter 7 Radial Symmetry Transform Based Eye Location

While the success of analytic face recognition approaches depends on the reliable detection of facial features, holistic approaches also need to use those feature points as important references for scale and orientation normalization. Eyes have been considered as more salient and stable than all other facial features (Brunelli et al., 1993). Face images can be easily normalized using geometrical measurements if both eyes are detected. Thus, eye location algorithms are very important for face recognition systems.

While the face recognition algorithms presented in previous chapters used manually located eye centres for normalization, a simple and robust eye location system with no training and extra device requirement is presented in this chapter. The approach is based on the generalized symmetry transform; a low level operator that can be applied successfully for detecting regions of interest without any priori knowledge (Reisfeld, Wolfson, & Yeshurun, 1995). Based on context free and low level components, a high level and purposive model, which utilizes prior knowledge of eye features, is then implemented for the eye location task. The performance of the algorithm has been fully tested using the BioID and BANCA database, and has also been integrated into an automatic face verification system.

7.1 Background

In (Yuille, Hallinan, & Cohen, 1992), deformable templates are used to detect facial features. The eye feature is described as a parameterized template which interacts dynamically with the image by altering its parameter values to minimize a defined energy function, thereby deforming itself to find the best fit. However, using this technique the templates have to be initialized at a position near to the actual eye location. The eigenface approach, was further developed in (Moghaddam & Pentland, 1994) in the form of eigeneyes, eigennoses and eigenmouths which were used to detect facial features. A Support Vector Machine approach is applied in (Huang, Shao, & Wechsler, 1998) to estimate the facial pose and detect the eye locations. 186 eye images and 186 non-eye image are used to train the SVM classifier. Both methods require many images for classifier and model training. Rizon et al. (Rizon & Kawaguchi, 2000) used intensity and edge information to detect candidates for facial features and a cost function is defined for each pair of feature points satisfying a spatial constraint. The pair of feature points with the smallest cost is determined to be the pupils of both eyes. A very different system was developed by Morimoto et al. (Morimoto, Koons, Amir, & Flickner, 2000) and applied to pupil detection. Two near infrared, time multiplexed light sources are synchronized with the camera frame rate to generate bright and dark pupil images, which are then used for pupil segmentation. However, most of the described algorithms were tested using only a small set of images, and their effects on the performance of face recognition/verification systems are seldom reported.

7.2 The Methodology

7.2.1 The Generalized Symmetry Transform

Since natural and artificial objects often give rise to the human sensation of symmetry, it has been suggested as one of the fundamental properties to guide higher level

processes in computer vision (Reisfeld et al., 1995). An object is regarded symmetric if it is invariant to the application of certain symmetry operations, e.g., the reflectional (mirror) symmetry operation. However, the shape of the object needs to be known before such operations can occur. The generalized symmetry transform, however, does not require this knowledge of shape. It operates on the edges in an image and assigns a continuous symmetry measure to each pixel.

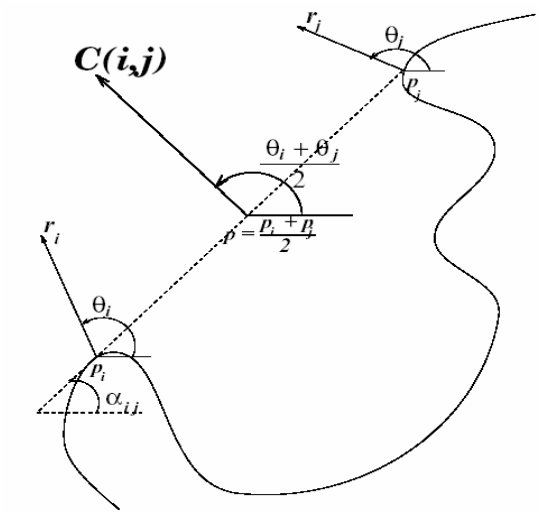


Figure 7-1 The contribution of points p_i and p_j to the symmetry measure (Reisfeld et al., 1995)

Let $p_k = (x, y)\ \ k = 1...K$ be any pixel in an image, and denote by $\nabla_{p_k} = (G_x, G_y)$ the horizontal and vertical gradient of the image at pixel p_k, i.e. $G_x = \partial p_k / \partial x, G_y = \partial p_k / \partial y$. The strength and phase of the gradient at p_k can be calculated as:

$$r_k = \log\left(1 + \left\|\nabla_{p_k}\right\|\right) \tag{7.1}$$

$$\theta_k = \arctan G_y / G_x \tag{7.2}$$

For each of the two points p_i and p_j, we define l as the line passing through them, with α_{ij} being the counter clockwise angle between l and the horizon. The direction of

symmetry axis for points p_i and p_j can be denoted as: $\varphi(p_i, p_j) = \frac{\theta_i + \theta_j}{2}$. We now also define a distance weight function and a phase weight function as:

$$D(p_i, p_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{[p_i - p_j]}{2\sigma}} \tag{7.3}$$

$$P(\theta_i, \theta_j) = (1 - \cos(\theta_i + \theta_j - 2\alpha_{ij}))(1 - \cos(\theta_i - \theta_j)) \tag{7.4}$$

The contribution of points p_i and p_j to the symmetry measure of point p can be represented as $C(i, j) = D(p_i, p_j)P(\theta_i, \theta_j)r_i r_j$. Now the symmetry magnitude of any point p can be defined as:

$$M(p) = \sum_{(p_i, p_j) \in \Gamma(p)} C(i, j)_j \tag{7.5}$$

where $\Gamma(p)$ is a set of points satisfying: $\Gamma(p) = \left\{ (p_i, p_j) \middle| \frac{p_i + p_j}{2} = p \right\}$. The symmetry magnitude thus averages the symmetry value over all orientations. Once the symmetry direction is defined as $\phi(p) = \varphi(p_i, p_j)$ such that $C(i, j) = D(p_i, p_j)P(\theta_i, \theta_j)r_i r_j$ is maximal for $(p_i, p_j) \in \Gamma(p)$, the symmetry value at point p can be denoted as:

$$S(p) = (M(p), \phi(p)) \tag{7.6}$$

7.2.2 The Radial Symmetry Measure

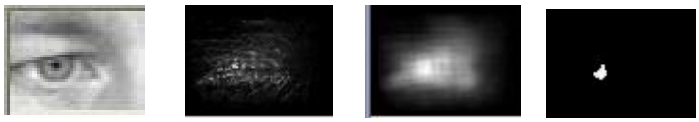
The transform defined above can effectively detect reflectional symmetry, which is invariant under 2D rotation and translation transforms. Sometimes we may also need to detect objects that are symmetric in multiple distinct orientations rather than a single principle one. The iris is an example of such an object. Radial symmetry such as this can be defined as:

$$RS(p) = \sum_{(p_i, p_j) \in \Gamma(p)} D(p_i, p_j)P(\theta_i, \theta_j)r_i r_j \sin^2(\varphi(p_i, p_j) - \phi(p)) \tag{7.7}$$

This expression emphasizes contributions in the orientations that are perpendicular to the main symmetry direction, and attains its maximum in a point that is surrounded by edges.

7.2.3 Eye Location by The Radial Symmetry

Since the main characteristic of the eye is its iris, which is symmetric in multiple distinct orientations, radial symmetry is adopted as our strategy for eye location. Our system incorporates the following modules: pre-processing, radial symmetry transformation, post-processing and eye location. Figure 7-2 shows the output from different modules within the system.



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

Figure 7-2 The system output at different stages. (a) the input image; (b) the radial symmetry map; (c) the filtered symmetry map; (d) the thresholded binary symmetry map

7.2.3.1 Input and Pre-processing

Once the face area is detected using a face detection module such as (Lienhart et al., 2002), the left and right eye regions can be roughly cropped and used as input to the system for precise eye centre location. To cope with variations caused by image noise and lighting, a 5×5 Gaussian filter is applied before the symmetry transform. This has proved to be a simple and effective solution for noise removal.

7.2.3.2 Radial Symmetry Transform and Post-processing

A symmetry magnitude map can be attained after applying the radial symmetry transform to the extracted eye region, as shown in Figure 7-2. From this figure one can see that the eye region has been highlighted. A 5×5 mean filter is then applied to the symmetry map for noise suppression. This is followed by a thresholding process in

which the symmetry map is now turned to a binary image, where the pixels with high symmetry values are assigned with the label '1' whilst the rest are assigned with label '0'. See Figure 7-2 for an example.

7.2.3.3 Eye Centre Location

The potential positions for the centre of each eye have now been reduced to several candidate areas, or as is true for most cases, a single region for each eye. Thus, the eye centre can now be trivially identified by locating the centre of each of the candidate areas. In cases where multiple candidate regions are still available, the smaller candidate regions are rejected. In addition, the candidate positions for both eyes are examined to ensure that the two eyes are located on, approximately, the same horizontal line. Figure 7-3 shows a sample face image and the eye centre locations extracted by the described algorithm.. More eye location results can be found in Figure 7-4.

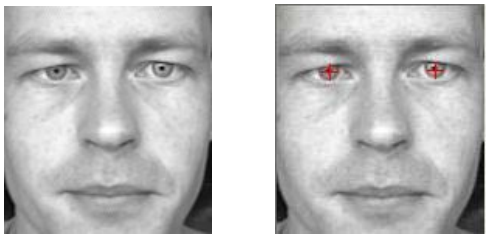


Figure 7-3 A sample face image and the located eye centre

7.3 Experimental Results

7.3.1 The Results on BioID Database

A test set, the BioID database (Jesorsky, Kirchberg, & Frichholz, 2001), is used in the experiments to evaluate the proposed algorithms. The set consists of 1521 images of 23 different people and was recorded during several sessions in multiple locations. This set features a large variety of illumination, background and face sizes. All of the images are grey scale images with a size of 384×288 pixels. The x and y coordinates of the left and right eyes are already indicated and recorded in text files. 1460 face images containing

prominent eyes are further selected from the database for additional testing. The test set thus contains 475 images captured from subjects with glasses and 985 images from subjects without glasses. Since the objective here is the fair evaluation of the eye location algorithms, the face area for each image is simply cropped according to the anthropometric relations between the face and facial features. Figure 7-4 shows a number of BioID face images and the eye centre locations extracted by the described algorithm (all images have been scaled to the same size for visual convenience).

Figure 7-4 Some sample test results

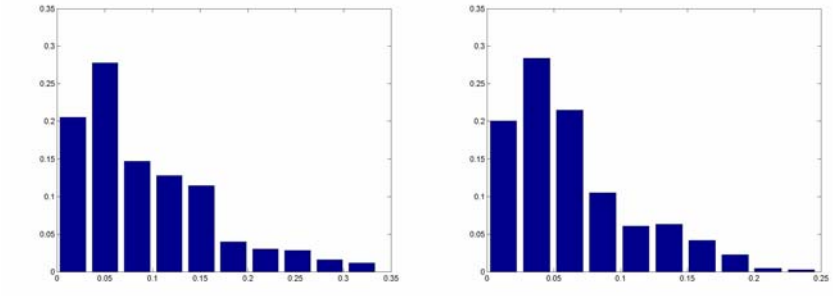


Figure 7-5 Error distribution for test set with glasses and without glasses

To evaluate the accuracy of the eye location algorithm, the normalized distance between the located eye centre (x_t, y_t) and the ground truth (x_c, y_c) is calculated as below:

$$d_e = \frac{\sqrt{(x_t - x_c)^2 + (y_t - y_c)^2}}{w} \tag{7.8}$$

where w is the distance between the ground truth left and right eye centres. A correct location of the eye in a face image is registered if the distance d_e is less than a threshold a, i.e., $d_e < a$. Figure 7-5 shows distribution of the error distance d_e for the test sets, both with and without glasses. One can observe that more than 91% of both histograms fall within the 0.2 error distance. The location accuracy for different values of a is shown in Figure 7-6, suggesting that 99.34% accuracy can be achieved for images without glasses when $a = 0.2$. Due to reflection and edges artifacts caused by wearing glasses, the figure drops to 91.26% for this test set.

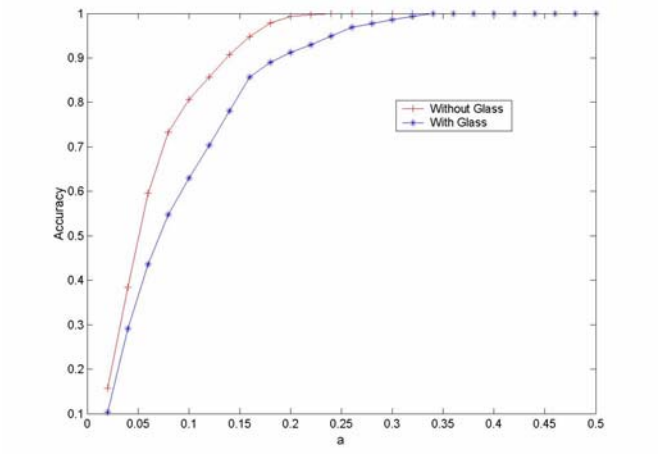


Figure 7-6 The location accuracy varying with parameter *a*

7.3.2 The Results on BANCA Database

After the automatic eye location algorithm is fully tested using face images from the BioID database, it is now integrated with an automatic face detection module (Lienhart et al., 2002) and tested using the BANCA database. The face detection module is

implemented as a cascade of Haar-like feature based classifiers, which have been shown to achieve a very good trade off between accuracy and detection efficiency (Viola et al., 2001). 2730 images from the development set of the BANCA database (Baillere et al., 2003) are used for testing. Sample images can be found in Figure 5-1. The test images are initially used as input into the face detection module to locate facial regions, on which the automatic eye location algorithm is then applied. Once the two eye centres are located they are used as reference points to enable the face images to be normalized in both rotation and scale.

Figure 7-7 Automatically normalized face images

Figure 7-7 shows some sample images, normalized using the automatic face and eye location system thus proving its robustness in a variety of situations. The sample images are captured in several different sessions: high quality camera with normal poses, low quality camera with normal poses and high quality camera with the head looking down are shown on the first, second and third rows respectively. There are, however, several cases where incorrect regions are selected due to errors in the face detection or eye location modules. For example, the images shown in Figure 7-8(a) are mainly background blocks, generated by errors in the face detection module. As shown in Figure 7-8 (b), the majority of the false "eyes" are located within hair regions, where many edges exist. Inclusion of too much background in the face region may also lead to

errors, since this can lead to an inaccurate initial guess for the eye regions. A statistical analysis of the eye location results on the BANCA database has also been performed and is shown in Table 7-1. In this analysis, the correct location of eyes is registered only if both eye centres are close enough to the ground truth data, in this case: $d_e < a, a = 0.2$.

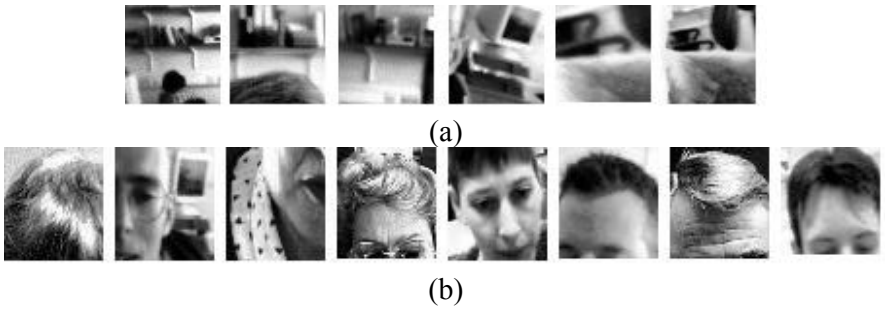


Figure 7-8 Wrong locations caused by face detection module (a); eye location module (b)

Number of test images	Performance of face detection module		Performance of eye location module	
	true face detections	false face alarms	true eyes locations	false eyes alarms
2730	2711 (99.30%)	15 (0.55%)	2674 (98.64%)	37 (1.36%)

Table 7-1 Statistical results on the BANCA database

7.3.3 Integration with the Face Verification System

Recall that a Gabor wavelet based face verification module has been developed in chapter 5. The verification system uses Gabor wavelets for feature extraction, GDA for enhancement and KNN for classification. The system is fully tested using the BANCA database according to the face verification competition held in 2004 (Messer et al., 2004). The comparative results show that the performance of the system is among the top methods. Since the system in chapter 5 normalizes images with manually located eyes, it is regarded as partially automatic. In the following experiments, an automatic verification system (Gabor + GDA) is developed by integrating the eye location module and tested using the same database. The results for the automatic verification system, together with that of other automatic algorithms, are shown in Table 7-2. The results

verify the robustness of both eye location and face verification algorithms proposed in this thesis. The performance of the Gabor + GDA method ranked within the top three and is significantly better than many other methods. The average error rate of Gabor + GDA is only 6.58%. Similar to the results reported in chapter 5, since a subject specific threshold is used, the other method Gabor + PCA + SVM developed by us achieves better performance than Gabor + GDA.

	R=0.1(WER)		***R=1(WER)***		***R=10(WER)***		**Avg**
	Dev	**Eval**	**Dev.**	**Eval.**	**Dev.**	**Eval.**	
IDIAP HMM	8.16	8.57	22.97	18.54	5.91	5.34	11.58
IDIAP Fusion	7.86	8.68	23.40	17.64	5.79	5.50	11.48
QUT	9.01	8.53	18.52	15.71	6.12	5.51	10.56
Univ Nottingham Gabor + PCA + SVM + subject specific thresholds	**3.34**	**3.20**	**8.51**	**7.59**	**2.51**	**2.30**	**4.58**
UniS-Fusion	7.92	10.06	16.07	18.00	4.58	5.42	10.34
UCL-LDA	9.77	10.84	20.30	19.55	7.21	6.97	12.44
UCL-Fusion	6.66	8.62	14.00	17.68	5.78	5.21	9.66
NeuroInformatik	34.00	36.00	16.70	16.10	36.80	37.70	29.55
Tsinghua Univ	**2.68**	**1.37**	**4.07**	**2.08**	**1.65**	**1.41**	**2.21**
CMU	7.72	8.78	26.08	23.05	18.19	9.12	15.49
***Gabor + GDA +* global threshold**	***5.37***	***6.73***	***10.22***	***10.08***	***3.73***	***3.36***	***6.58***

Table 7-2 Verification results for fully automatic systems

Due to minor errors within the automatic eye location algorithms, the automatic verification algorithms normally exhibit higher levels of verification error than the partially automatic methods. The comparative results of the fully automatic system with the corresponding partially automatic system, which normalize faces with manually located eyes, are shown in Table 7-3. The results for several different research institutions on the same BANCA database are also included, see (Messer et al., 2004) for the details. As expected, the automatic face detection and eye location module increase the weighted error rate of Gabor + GDA from 4.48% to 6.58%, which is common for all of the verification systems. However, the performance of the eye

location system is state of the art, and the developed automatic verification system achieves significantly lower error rates than many other systems.

	Error Rate with Manual Eye Location (%)	*Error Rate with Automatic Eye Location (%)*	*Increase of Error Rate (%)*
Gabor + GDA + global threshold	4.48	6.58	46.87
Gabor + PCA + SVM + subject specific thresholds	3.33	4.58	37.54
CMU	8.11	15.49	90.99
Univ. of Surrey	7.99	10.34	29.41
Tsinghua Univ.	1.39	2.21	58.99

Table 7-3 Comparative results for fully and partially automatic face verification systems

7.4 Conclusions

A generalized symmetry transform based eye location algorithm has been proposed in this chapter. The robustness of the algorithm is first tested using 1460 face images from the BioID database, 99% and 93% accuracy was achieved for face images with and without glasses respectively. The eye location algorithm has also been tested using 2730 images from the BANCA database, about 98.6% accuracy has been achieved. The results suggest that a more precise face locator could alleviate many of the eye location errors. An automatic verification system has been further developed by integrating the eye location module with the verification module (Gabor + GDA) proposed in chapter 5. The automatic system is fully tested using the BANCA database according to protocols used in the recent Face Verification Competition 2004. Though the error rate is larger than that reported in chapter 5 due to the mis-alignment among face images caused by the eye location algorithm, the performance of the automatic system is one of the top three and better than most of the participants in the contest.

Chapter 8 The Developed User Identification System

This chapter presents an automatic user identification system developed at the initial stages of this research. The system consists of the following modules: face detection, registration and user information management. Once a subject is registered with the system, it can identify the registered person in real time when his face image is detected from a web cam. Based on its efficiency, the system is further developed to identify multiple persons simultaneously from real video streams. A video demo displaying how the system works can be found at http://www.cs.nott.ac.uk/~lls/demo.htm.

8.1 System Architecture

8.1.1 Registration

Each candidate needs to be registered with the system before they can be identified. The registration process thus consists of the following modules: user information registry, face detection, feature extraction and/or model training and feature/model saving. As shown in Figure 8-1, the process requires the full support of the face detection module, user management module and recognition module, which will be described in detail in the next section. About 30 staff from the Nottingham Computer Science School are registered with our system, with at least 5 face images for each subject on record. Once the face images are registered, the recognition module can be invoked to extract features or train subject specific models. These are then saved via the user management module for future identification purposes.

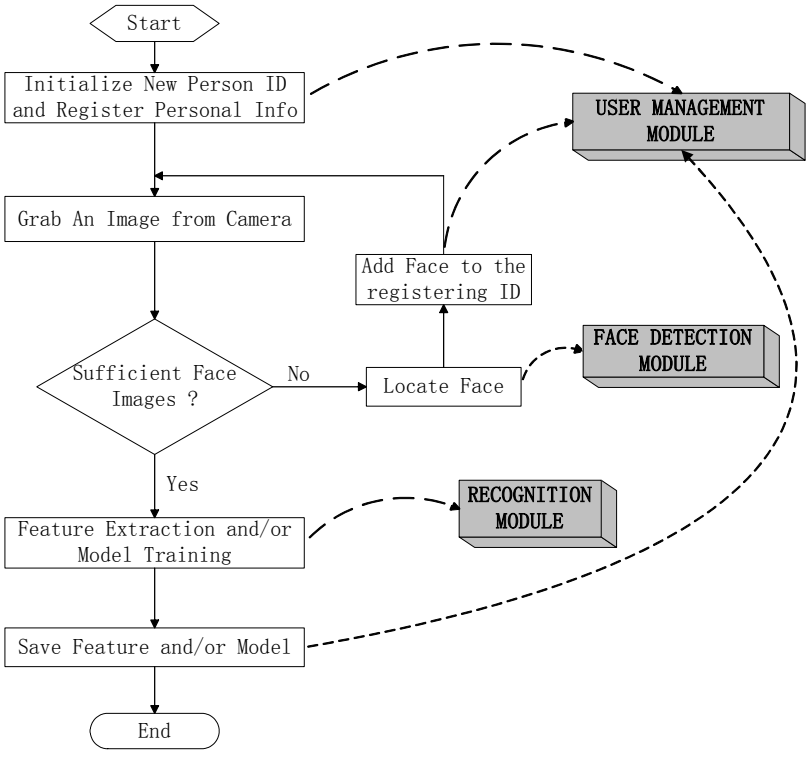


Figure 8-1 Registration flow chart

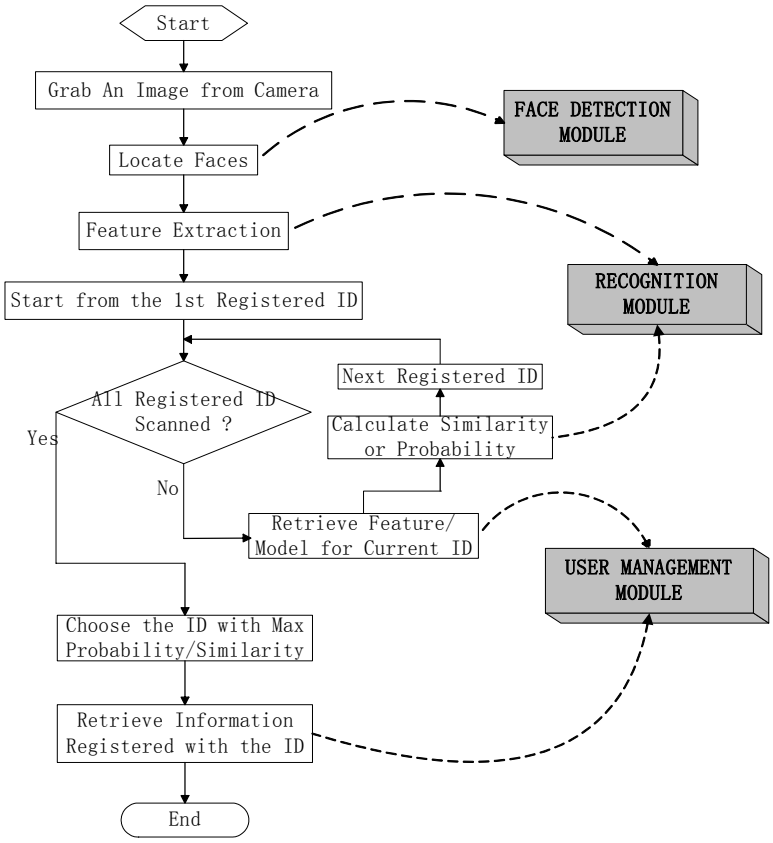


Figure 8-2 Identification flow chart

8.1.2 Identification

The aim of user identification is to identify a subjects ID when their face is presented before a web cam. The subject has to be registered with the system before he can be identified. The identification process, as shown in Figure 8-2, can be summarized as follows: when a user is sitting before the web cam, their face area is located and captured, then refined by the face detection module and finally passed to the recognition module for processing and identification. The recognition module compares the input face with each registered subject, either by matching features directly or by computing the probability. The face is then identified as the person whose features or model gives the maximum similarity or probability. The personal information registered with the ID will finally be retrieved from the user management module and presented

by the system. Figure 8-3 shows a snapshot of the system. The screenshot shows that the system correctly identify the user "Dylan Shen" and presents his personal information, i.e., name, address, age, etc. The left column in the interface shows a subset of the users who have had their faces registered with the system.

Figure 8-3 A snapshot of the user identification system

8.2 System Modules

8.2.1 Face Detection

The algorithm proposed in (Lienhart et al., 2002) is initially applied in the system for face detection. The method is a classification based algorithm, which cascades a series of Haar-like features based face/non-face classifiers for efficient detection. The set of classifiers are all trained using AdaBoost algorithm and combined to form the final classifier, more details can be found in (Viola et al., 2001; Lienhart et al., 2002). Once the efficient classifier is learned, a window will be used to scan the test image to search for face instances. Source code for the face detector is freely available at the Intel Open Source Computer Vision Library (Intel Corporation, 2005). Figure 8-4 shows a sample image with the located face marked in a red rectangle. As can be seen, the located face

area contains a lot of noise information, e.g. background and hair etc, which could affect the performance of recognition algorithm. A skin mask module is developed and integrated into the system in order to refine the results from the initial face detection process.

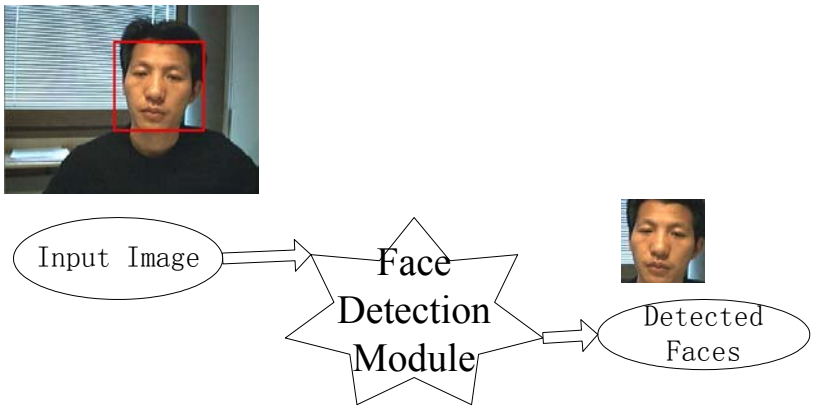


Figure 8-4 A sample image with detected face

8.2.1.1 Skin Masking

It is widely accepted that the colour of human skin is distinctive from the colour of many other natural objects. Analyzing the statistics on skin colour it can be observed that skin colours are distributed over a small area in the chrominance plane with the major difference between skin tones being variations in intensity (Menser & Muller, 1999). To utilize skin colour properties for the face detection refinement process an image is first converted into luminance and chrominance channels in the YCbCr color space.

Let $\mathbf{w}_{ij} = [Cb_{ij}\ Cr_{ij}]^T$ denote a vector composed of the chrominance components Cb and Cr for a pixel *(i, j)*. The class-conditional pdf of $\mathbf{w}_{ij}$ belonging to the skin class *x* is modeled by a two-dimensional Gaussian (Menser et al., 1999; Bai & Shen, 2003b):

$$p(\mathbf{w}_{ij} \mid x) = (2\pi)^{-1} \mid \Sigma \mid^{-1/2} \exp\left(-\frac{1}{2}\left[\mathbf{w}_{ij} - \boldsymbol{\mu}\right]^T \Sigma^{-1}\left[\mathbf{w}_{ij} - \boldsymbol{\mu}\right]\right) \qquad (8.1)$$

where the mean vector $\boldsymbol{\mu}$ and the covariance matrix Σ are estimated from the training set. Figure 8-5 shows the distribution of skin colours in the Cb and Cr domains. The contour of the pdf defines an ellipse in the CbCr domain, whose center and principal axis are determined by $\boldsymbol{\mu}$ and Σ, respectively. After building the skin colour model, the original colour image can be easily converted to a skin probability image *P* using equation (8.1). The image *P* indicates the probability of each image pixel belonging to the skin class *x*, i.e., $P(i,j) \sim p(\mathbf{w}_{ij} \mid x)$. Figure 8-6 (a) and (b) shows the input colour image *I* and the skin probability image *P* respectively.

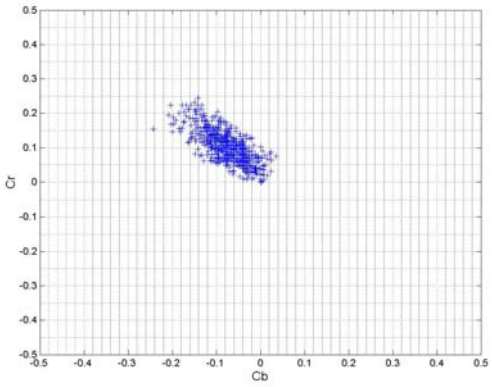


Figure 8-5 Distribution of skin colors in Cb, Cr domain

8.2.1.2 Ellipse Masking and Head Orientation Estimation

Once the face region is extracted from the input image, a ellipse fitting method (Bradski, 1998) can be used to approximate the skin blob and estimate the head orientation. Details of the fitting algorithm can also be found in Appendix C, which is based on statistical analysis of the skin probability image. Once the parameters of the ellipse approximation of the skin blob are determined, a face image can be masked by the ellipse with major axis *l*, minor axis *w*, and orientation θ. Figure 8-6 (c) shows the ellipse masked face image from (a). Figure 8-7 shows the fitted face with different

orientations. Each face has been masked with a corresponding ellipse. The two axes, centroid and orientation of the ellipse are indicated by a cross.

(a) (b) (c)

Figure 8-6 Detected face image (a); skin probability image (b) and masked face image (c)

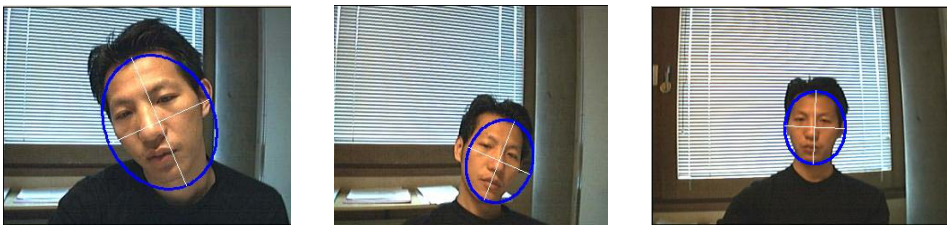


Figure 8-7 Ellipse fitting for faces with different orientations

8.2.2 Recognition

As shown in Figure 8-8, the recognition module works in two modes: registration and identification. While the module extracts features and/or trains models for future processing in registration mode, it must compare the test face with each registered ID when in identification mode. A HMM based face recognition method is adopted in this system, which treats a face image as a sequence of states produced when the face is scanned from top to bottom. More interesting is the 2D embedded HMM proposed by Nefian (Nefian et al., 1999). The embedded HMM consists of a set of super states with each super state being associated with a set of embedded states. Super states represent primary facial regions whilst embedded states within each super state describe in more detail the facial region. Nefian defined 5 super states: forehead, eyes, nose, mouth and chin. Transitions between embedded states in different super states are not allowed. In a HMM based face recognition implementation, a face image is divided into a series of overlapping image blocks, the observation sequence can then be generated by

concatenating the observation vectors extracted from each image block for HMM training. Once HMM models are trained using registered face images, the observation sequence extracted from a test image is used as input to all of the trained HMMs associated with each person and the conditional probability given by each HMM is calculated. The identity of the input face is determined by the HMM which produces the highest probability. Figure 8-9 shows the flow chart of a generic HMM based face recognition system.

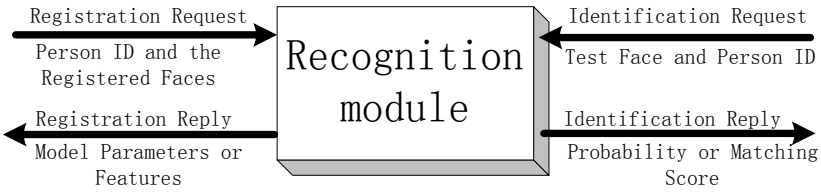


Figure 8-8 Recognition module diagram

The observation vectors O_t could be simply the grey values of pixels in the image block. However, such a method is sensitive to image variation due to illumination, translation and rotation. Moreover, since the dimension of the observation vectors is high, much computation is required. Image transform techniques will be helpful to make the model more robust and perform feature dimension reduction at the same time. Nefian et al apply 2D Discrete Cosine Transform (DCT) on each image block and only the low frequency coefficients are extracted to produce observation vectors. Due to its origins in simultaneous time and frequency analysis, wavelets are widely believed to be advantageous for image representation over other mathematical transforms such as the Fourier transform or DCT. Therefore, a Discrete Wavelet Transform (DWT) based HMM has also been proposed in (Bai et al., 2003a) for face recognition. Compared with DCT, DWT based HMMs achieved higher accuracy at the expense of slightly reduced

efficiency. Both methods have been implemented in the system and can be switched between according to application requirements.

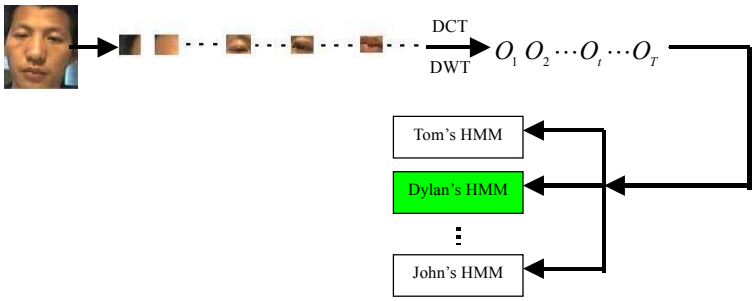


Figure 8-9 The HMM face recognition algorithm

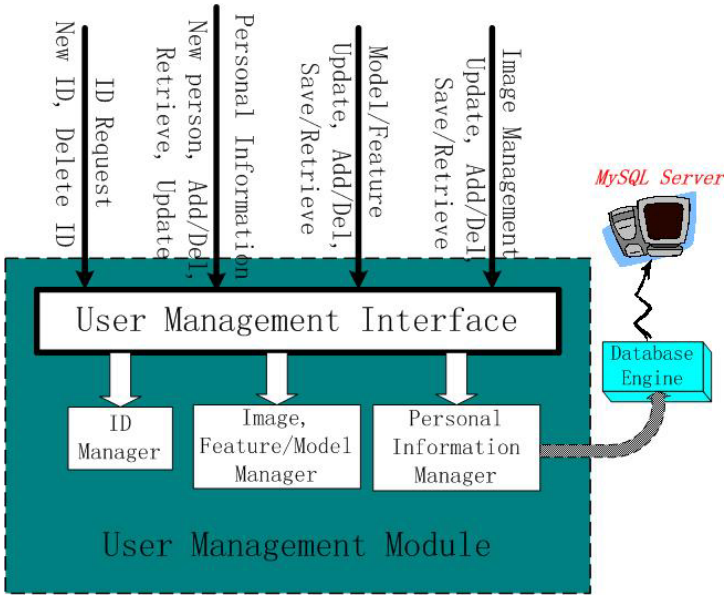


Figure 8-10 User management module diagram

8.2.3 User Management

ID management, image feature/model management and personal information management are the three main functions of the user management module (see Figure 8-10 for details). The image feature/model manager is mainly concerned with the recording of all image feature/model files for registered subjects. The record gives an overview of the face database, as well as the details about saved image feature/model files, e.g. the path, number etc. The module updates each record whenever there is a

relevant change and responds to image feature/module retrieval requests when queried with a user ID. The user ID manager is mainly responsible for the issuing of new IDs and removing old IDs. The personal information manager maintains data regarding each registered users name, address, age and sex etc. Since the data is stored on a MySQL server, the personal information manager requires a database engine to interpret the Add/Delete, Update and Query SQL requests.

8.3 Conclusions

An automatic face based user identification system has been presented in this chapter. When integrated with face detection, recognition and user management modules, the system can locate faces from images captured by a normal web cam and recognize a subjects identity in real time. The flowcharts for two of the most important processes (registration and identification) have been described and the main functions of the three system modules have been explained in detail. A database with about 30 subjects, who are mainly students and staff from the University of Nottingham Computer Science School, has also been built to test the system. The system has shown excellent performance with high efficiency when this small database is used. Based on this framework, a video based face identification system has also been developed. The system can detect multiple faces in a real time video stream and identify each of them. Figure 8-11 shows a snapshot of the video based system, where three faces are detected, identified and labelled with the registered names. A demo of the system can also be found at: http://www.cs.nott.ac.uk/~lls/demo.htm. The system, when running on a P4-1.8GHz PC, can support video streams with frame rates of up to 3 frames/sec.

Figure 8-11 A snapshot of the video based identification system

Since the system was developed at the initial stages of the research, the HMM based recognition algorithm is adopted. However, the algorithm has been shown to be only suitable for small databases. The results reported in chapter 4 show that though the DWT-HMM method achieves 97.5% accuracy at the ORL database (40 subjects), the figure drops dramatically to 44.5% on the subset of the FERET database (200 subjects) used for testing. A more robust method, such as the described Gabor wavelet based approach, which has been fully tested using a number of large databases in this thesis, could easily be integrated into the system framework for additional performance improvements.

Chapter 9 Conclusions and Future Works

A fast and robust Gabor wavelet based method has been proposed for face recognition in this thesis and the method has been fully tested using public databases, e.g. FERET, BANCA etc. This chapter will give a summary about the work presented in previous chapters and some suggestions for future developments.

9.1 Summary of Works

9.1.1 An Overview of Gabor Wavelets: Background and Applications

A detailed review of the background and applications of Gabor wavelets has been presented in this thesis. Contributed by Dennis Gabor in 1946, the 1D Gabor function was first proposed for joint time frequency analysis of the time signal. As a member of the wavelet family, mathematical analysis shows that the Gabor wavelet achieves the optimal resolution in both the time and frequency domains. In the spatial domain, researchers have presented evidence showing the similarity of 2D Gabor wavelets with the receptive fields of mammalian visual cortex cells. Motivated by the mathematical background and biological evidence, 2D Gabor wavelets have been widely applied in different computer vision and pattern recognition applications including face recognition. A literature review on the application of Gabor wavelets for face representation has also been performed in this research. Aiming to give some guidance to researchers in this area, the review presented the latest Gabor wavelet based methods available in the literature and discussed both the limitations and advantages of different approaches.

9.1.2 Gabor Wavelets and Kernel Subspace Methods for Face Identification and Verification

Though face recognition has been an active research area for many years, it is still an unsolved problem due to the complex distortions caused by expression, pose and illumination variation. However, the task seems to be trivial for human beings. With the aid of complex perceptual systems, such as the visual cortex, it is very common for a human to recognize thousands of people, even in the presence of dynamic variations of face shape, pose, expression and appearance. Based on the overview on background and applications of Gabor wavelets, they are adopted in this research as a method to extract

robust features for face recognition purposes. Once the features are extracted, nonlinear kernel subspace analysis, i.e. GDA, is further applied for dimension reduction and class separability enhancement. The combination of Gabor wavelets and kernel methods have been successfully applied to face identification and verification and fully tested using public databases, e.g. ORL, FERET and BANCA. While the proposed method has achieved better performance than other state of the art identification algorithms on the ORL and FERET database, it has also shown to be more robust than most of the participants in the recent face authentication test using the BANCA database.

9.1.3 Learning the Most Important Gabor Features for Object Detection and Recognition

Despite the robustness of Gabor wavelets based methods, they require high computation and memory cost. Since a set of 40 wavelets is convolved with images, the feature extraction process takes long time. Though FFT could be used to speed up the convolution process, the huge dimension of extracted features will also bring high computation cost to the classification process. As a result, a feature selection method is required to eliminate those redundant features for dimension reduction. In this thesis, the AdaBoost algorithm is first applied to select Gabor features for object detection. Since both feature selection and classifier training can be completed in the same learning process, the classifier using the selected Gabor features can be used for object detection directly. A novel feature selection algorithm, MutualBoost, has also been proposed and successfully applied to select Gabor features for face recognition. Particularly, the mutual information between candidate features is used as an additional criterion to select one by one the most important Gabor features. Compared with AdaBoost selected features, the results show that Gabor features learned using MutualBoost techniques are more discriminative and achieve better recognition

accuracy. Both systems have been compared with those using the pre-selected Gabor features, substantial efficiency improvements have been observed without performance deterioration. The face recognition system using the selected Gabor features has also been compared with other state of the art methods on the whole FERET database according to the evaluation protocol and better accuracy has been achieved. The face recognition system thus developed is both robust and efficient.

9.1.4 Automatic Eye Location

To normalize the scale and orientation of different face images, an automatic eye location algorithm is required before the robust Gabor feature based face recognition system can be applied in real applications. Though there are quite a number of complex methods available, most of the eye location systems are only tested using a limited number of images and they normally require lots of training samples. The method proposed in this thesis is, however, very simple and requires no training images. The approach is based on a context free feature detector, the generalized symmetry transform, which requires no prior knowledge about eyes. Once those areas with large symmetry values are located, eyes can be easily located at the centre of these regions. The robustness of the algorithm is first tested using 1460 face images from the BioID database, 99% and 93% accuracy are achieved for face images with and without glasses respectively. The eye location algorithm has also been tested using 2730 images from the BANCA database, about 98.6% accuracy has been achieved. Based on the proposed eye location module, a fully automatic verification system has also been developed by integrating the verification module (Gabor + GDA) proposed in chapter 5. The automatic system is tested using the BANCA database according to protocols defined by the recent Face Verification Competition 2004. The performance of the automatic

verification system is one of the top three and better than most of the participants in the contest.

9.1.5 The User Identification System

An automatic real time user identification system has been developed in this research. The system consists of three main modules: face detection, recognition and user management. With the full support of each of these modules, the system can efficiently detect faces from images captured by a web cam, extract features and identify the user. The system can also function in registration mode such that the personal information, face images and model/features can be registered and saved either in files, or in the MySQL database. Utilising the high efficiency of the proposed techniques, a video based face identification system has been further developed, which can detect multiple faces from a real time video stream, identify them and display their names. The modular design of the system allows a large degree of flexibility, allowing for future expansion and the integration of any new face detection or recognition algorithms.

9.2 Future Works

9.2.1 Extensions of the Present Works

9.2.1.1 A Complete Gabor Feature Based Object Detection system

Though the Gabor feature based classifier has shown the ability to discriminate car and non-car images as well as face and non-face images, more works still need to be done before the classifier can be applied in real object detection applications. For classification based detection methods, an image is usually scanned by a $n \times n$ window with one pixel step size. Each image window is then input to the learned classifier to make a classification decision, i.e., object or background. Figure 9-1 shows a typical classification based face detection system. To deal with the scale variance, the image is usually rescaled by s different factors such that a set of multi-resolution images are

generated, the detection process can be applied to each image thereafter. As a result, the number of images to be processed by the classifier is huge, with more than 90,000 windows ($n = 20$) needing to be classified for an image with size 200×200 when $s = 5$.

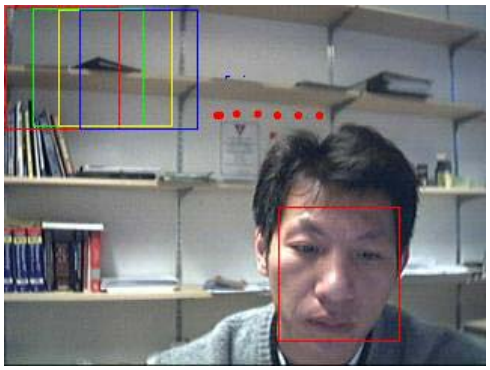


Figure 9-1 A classification based face detection system

Based on the fact that most of the scanned image blocks are actually background (see Figure 9-1), a cascade of classifiers is used in (Lienhart et al., 2002; Viola et al., 2001) to speed up the detection process. Figure 9-2 shows the cascade structure of three classifiers. Simple classifiers are used to reject the majority of the sub windows before more complex classifiers are applied. The simple classifiers are adjusted such that the false negative rate is close to zero. A positive result from the first classifier triggers the evaluation of the second classifier with high detection rates, and so on. A negative result at any point leads to the immediate rejection of the sub window. As such, the cascade attempts to reject as many negative windows as possible at the earliest stage possible. Such a cascade structure shall also be used to learn a Gabor feature based classifier for real time object detection. The classifier at the 1st stage could be one which uses only two Gabor features with a minimized false negative rate. Subsequent classifiers will require a larger number of features. To reduce the computation cost of feature extraction, the classifiers at early stages could also be trained using simpler

features, e.g. Haar-like features (Lienhart et al., 2002; Viola et al., 2001). The complete object detector will thus be both robust and fast.

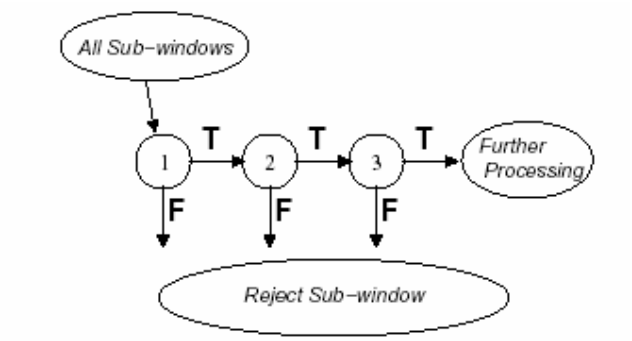


Figure 9-2 Diagram of the detection cascade (Viola et al., 2001)

9.2.1.2 Effects of Eye Location Algorithm on Face Recognition Approach

A fully automatic face verification system has been developed in this thesis using a generalized symmetry based eye location algorithm. The Gabor + GDA system combines Gabor features and GDA for verification. Though the error rate is higher than that of the approach using manually located eyes for normalization, the performance of the fully automatic system has shown to be much better than many state of the art methods. The test proves the robustness of Gabor features against the mis-alignment caused by automatic eye location algorithms. To reduce the complexity and memory cost of Gabor feature based methods, information theory has also been applied to select the most important features for recognition. The method, named as MutualGabor + GDA, has achieved better recognition performance than Gabor + GDA when manually located eyes are used for image normalization. Since Gabor features are extracted from local image regions and MutualGabor + GDA is actually using a subset of the features used by Gabor + GDA, the method should be as robust against mis-alignment as Gabor + GDA. However, future experiments should be carried out to justify this argument.

9.2.2 Gabor Feature Selection with Larger Search Space

Since a Gabor feature is simply extracted by applying a wavelet at a certain image location, the mutual information based method in this thesis selects both Gabor wavelets and the locations where the wavelets are applied. While Gabor wavelets with varied frequency and orientations are applied at different locations, the approach reflects the fact that different image regions display varied texture features. However, the candidate features in this work are extracted using a pre defined set of 40 Gabor wavelets. While the most appropriate wavelet in the candidate set is chosen for a certain image location, the optimal wavelet for the position might not be included in the defined set. The search space of the wavelets has to be extended to all possible parameter spaces such that the optimal one can be found. Two ratios between the Gaussian envelop sizes and the central frequency, orientation, image location and centre frequency now form the five dimensional parameter space of candidate Gabor wavelets. The parameter space to be searched will be significantly larger as more parameters are included. Whilst most optimization algorithms seem to be intractable in the context of this problem, genetic algorithms (GA) may prove to be a suitable choice. However, the computation burden has to be reduced before GA can be applied.

9.2.3 Pose Invariant Face Recognition

Motivated by the biological resemblance with the primary visual cortex, Gabor wavelets form an optimal basis for measuring local texture features and representing images. By their very nature, Gabor wavelet representations are to some extent insensitive to variations of lighting and local distortions caused by face position and expression. Extensive experiments have shown the success of Gabor features for frontal view face recognition in this thesis. However, faces are such complex patterns that many images captured from the dynamic real world are often half profile, or even full

profile. A pose estimation module could be used to decide the pose of test faces such that they are compared with the images with corresponding pose only. However, a lot of images crossing different poses would need to be saved in the database for each subject, which would significantly increase the applications memory cost requirements. A 3D model could also be used to synthesize frontal view images from faces with different poses, though the model fitting process takes long time. Switching to 3D range data could alleviate the pose problem since the depth data can be easily rotated in 3D space. Since 3D face recognition systems are still in the early development stages, the robustness and accuracy of such systems remains unclear. However, 3D face models or 3D face recognition systems should be involved in future works to develop fully pose invariant recognition systems.

Appendix A Eigenvalue Solutions of GDA

The method to solve the eigenvalue problem: $\lambda \mathbf{KK\alpha} = \mathbf{KUK\alpha}$ is shown in this appendix.

The solution starts with the eigen decomposition of the Kernel matrix $\mathbf{K}$:

$$\mathbf{K} = \mathbf{P\Gamma P}^{t} \qquad \text{(A.1)}$$

where $\mathbf{\Gamma}$ is a diagonal matrix consisting of the eigenvalues of $\mathbf{K}$ and $\mathbf{P}$ is the matrix of normalized eigenvectors associated with $\mathbf{\Gamma}$. Thus $\mathbf{\Gamma}^{-1}$ exists and $\mathbf{P}^{t}\mathbf{P} = \mathbf{I}$.

Substituting (A.1) to $\lambda \mathbf{KK\alpha} = \mathbf{KUK\alpha}$:

$$\lambda \mathbf{P\Gamma P}^{t}\mathbf{P\Gamma P}^{t}\boldsymbol{\alpha} = \mathbf{P\Gamma P}^{t}\mathbf{UP\Gamma P}^{t}\boldsymbol{\alpha} \qquad \text{(A.2)}$$

Multiply $(\mathbf{P\Gamma})^{-1}$ on both sides and given $\mathbf{P}^{t}\mathbf{P} = \mathbf{I}$, we obtain:

$$\lambda \mathbf{\Gamma P}^{t}\boldsymbol{\alpha} = \mathbf{P}^{t}\mathbf{UP\Gamma P}^{t}\boldsymbol{\alpha} \qquad \text{(A.3)}$$

Now define $\boldsymbol{\beta}$, $\mathbf{V}$ such that $\boldsymbol{\beta} = \mathbf{\Gamma P}^{t}\boldsymbol{\alpha}$ and $\mathbf{V} = \mathbf{P}^{t}\mathbf{UP}$, (A.3) can be simplified as:

$$\lambda \boldsymbol{\beta} = \mathbf{V}\boldsymbol{\beta} \qquad \text{(A.4)}$$

Once $\boldsymbol{\beta}$ is calculated using the eigen decomposition of matrix $\mathbf{V}$, $\boldsymbol{\alpha}$ can be simply obtained using:

$$\boldsymbol{\alpha} = \mathbf{P\Gamma}^{-1}\boldsymbol{\beta} \qquad \text{(A.5)}$$

Appendix B Optimising α_t and h_t in AdaBoost Algorithm

This appendix shows how to find the value of α_t and h_t to minimize the training error $\varepsilon_{tr}(H)$ of the learned strong classifier H by the AdaBoost algorithm.

As shown in chapter 3, the training error is bounded by $\varepsilon_{tr}(H) \le \prod_t Z_t$, where $Z_t = \sum_i w_t(i)\exp(-\alpha_t y_i h_t(x_i))$. To simplify the notation, let us fix t and let $u_i = y_i h_t(x_i)$, $Z = Z_t$, $w = w_t$, $h = h_t$ and $\alpha = \alpha_t$. Our goal is to find the value α which minimizes or approximately minimizes Z as a function of α.

For weak hypotheses h with range $[-1,1]$, Z can be approximated as follows:

$$\begin{aligned} Z &= \sum_i w(i)\exp(-\alpha u_i)) \\ &\le \sum_i w(i)\left(\frac{1+u_i}{2}\exp(-\alpha) + \frac{1-u_i}{2}\exp(\alpha)\right) \end{aligned} \tag{B.1}$$

This upper bound is in fact exact if h has range $\{-1,1\}$, which can be further rewritten as:

$$Z = \frac{\exp(-\alpha)}{2}\left(1 + \sum_i w(i)u_i\right) + \frac{\exp(\alpha)}{2}\left(1 - \sum_i w(i)u_i\right) \tag{B.2}$$

Solving the equation $\partial Z / \partial \alpha = 0$, we can find the optimal value of α to be:

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1+\sum_i w(i)u_i}{1-\sum_i w(i)u_i}\right) \tag{B.3}$$

Plugging into Eq.(B.2), this setting gives the upper bound $Z \le \sqrt{1-r^2}$, where

$$r = \sum_i w(i)u_i = \sum_i w(i)y_i h(x_i) \tag{B.4}$$

The quantity r is actually a natural measure of the correlation of the predictions of h and the labels y with respect to the distribution w. Let $[\pi]$ be an indicator variable that is 1 if the predicate π is true and 0 otherwise, r can be related with the weighted error $\varepsilon = \sum_i w(i)[h(x_i) \neq y_i]$ of h as follows:

$$\begin{cases} \sum_i w(i)[h(x_i) = y_i] - \sum_i w(i)[h(x_i) \neq y_i] = r \\ \sum_i w(i)[h(x_i) = y_i] + \sum_i w(i)[h(x_i) \neq y_i] = 1 \end{cases} \quad \text{(B.5)}$$

Solving these equations, we get:

$$\varepsilon = \sum_i w(i)[h(x_i) \neq y_i] = \frac{1-r}{2} \quad \text{(B.6)}$$

Once the setting of α is found, the weights of samples can be updated as follows:

$$w_{t+1}(i) = \frac{w(i)\exp(-\alpha u_i)}{Z} \quad \text{(B.7)}$$

Recall that α is selected such that Z is minimized, i.e. $\partial Z / \partial \alpha = 0$. By definition

$$\begin{aligned} \frac{\partial Z}{\partial \alpha} &= \frac{\partial\left(\sum_i w(i)\exp(-\alpha u_i))\right)}{\partial \alpha} \\ &= -\sum_i w(i) u_i \exp(-\alpha u_i)) \\ &= -Z\sum_i w_{t+1}(i) u_i \end{aligned}, \quad \text{(B.8)}$$

Thus, $\sum_i w_{t+1}(i)u_i = \sum_i w_{t+1}(i) y_i h(x_i) = 0$. In other words, this means that, with respect to the updated distribution w_{t+1}, the prediction of h selected in round t will now be exactly uncorrelated with the labels y_i.

Appendix C Skin Blob Ellipse Fitting

Assume a face image *I(x,y)* with size *W*×*H* is detected by the face detection module, and *P(x,y)* is the skin probability value for pixel *(x,y)*. The zeroth moment of the skin probability image *P* is:

$$M_{00} = \sum_{x=1}^{W}\sum_{y=1}^{H} P(x,y), \tag{C.1}$$

the first moments are:

$$M_{10} = \sum_{x=1}^{W}\sum_{y=1}^{H} xP(x,y), M_{01} = \sum_{x=1}^{W}\sum_{y=1}^{H} yP(x,y), M_{11} = \sum_{x=1}^{W}\sum_{y=1}^{H} xyP(x,y), \tag{C.2}$$

and the second moments are defined as:

$$M_{20} = \sum_{x=1}^{W}\sum_{y=1}^{H} x^2 P(x,y), M_{02} = \sum_{x=1}^{W}\sum_{y=1}^{H} y^2 P(x,y) \tag{C.3}$$

Then the centre position (x_c, y_c) of the skin blob in the face region is:

$$x_c = \frac{M_{10}}{M_{00}}, y_c = \frac{M_{01}}{M_{00}} \quad , \tag{C.4}$$

and the head orientation can be defined as:

$$\theta = \frac{arcta\left(\dfrac{2\left(\dfrac{M_{11}}{M_{00}} - x_c y_c\right)}{\left(\dfrac{M_{20}}{M_{00}} - x_c^{\,2}\right) - \left(\dfrac{M_{02}}{M_{00}} - y_c^{\,2}\right)}\right)}{2} \tag{C.5}$$

The length *l* and width *w* of the skin blob can be calculated as follows:

$$l = \sqrt{\frac{(a+c) + \sqrt{b^2 + (a-c)^2}}{2}}, w = \sqrt{\frac{(a+c) - \sqrt{b^2 + (a-c)^2}}{2}} \tag{C.6}$$

where

$$a = \frac{M_{20}}{M_{00}} - x_c^{\,2}, b = 2\left(\frac{M_{11}}{M_{00}} - x_c y_c\right), c = \frac{M_{02}}{M_{00}} - y_c^{\,2} \quad \text{(C.7)}$$

Bibliography

Agarwal, S., Awan, A., & Roth, D. (2002). The UIUC image database for car detection. http://l2r.cs.uiuc.edu/~cogcomp/Data/Car/ [On-line].

Aslam, J. (2000). Improving Algorithms for Boosting. In *Proceedings of the Thirteenth Annual Conference on Computational Learning Theory* (pp. 200-207).

Ayinde, O. & Yang, Y. H. (2002). Face recognition approach based on rank correlation of Gabor- filtered images. *Pattern Recognition, 35,* 1275-1289.

Bai, L. & Shen, L. (2003a). Combining wavelet and HMM for face recognition. In *Proc. Of the 23rd Artificial Intelligence Conference* (pp. 227-234). Cambridge,UK.

Bai, L. & Shen, L. (2003b). Face detection by orientation map matching. In *Proc. of International Conference on Computational Intelligence for Modelling Control and Automation* (pp. 363-370). Vienna, Austria.

Baillere, E. B. & Bengio, S. (2003). The BANCA database and evaluation protocol. In *Proc. of the 4th International Conference on Audio- and Video-Based Biometric Person Authentication* (pp. 625-638). Guilford, UK.

Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Transactions on Neural Networks, 13,* 1450-1464.

Baudat, G. & Anouar, F. E. (2000). Generalized discriminant analysis using a kernel approach. *Neural Computation, 12,* 2385-2404.

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

Beumier, C. & Acheroy, M. (2000). Automatic 3D face authentication. *Image and Vision Computing, 18,* 315-321.

Boser, B. E., Guyon, I. M., & Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In *Proc. of the 5th Annual Workshop on Computational Learning Theory* (pp. 144-152). Pittsburgh.

Bowyer, K. W., Chang, K., & Flynn, P. (2004). A survey of 3D and multi-modal 3D+2D face recognition. In *Proc. of International Conference on Pattern Recognition* (pp. 358-361). Cambridge,UK.

Bradski, G. (1998). Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal, 2,* 1-15.

Brunelli, R. & Poggio, T. (1993). Face Recognition - Features Versus Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15,* 1042-1052.

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, 2,* 121-167.

Carbonetto, P. (2001). Viola training data. http://www.cs.ubc.ca/~pcarbo/viola-traindata.tar.gz [On-line].

Chellapa, R., Wilson, C., & Sirohey, S. (1995). Human and machine recognition of faces: a survey. *Proceedings of the IEEE, 83,* 705-740.

Chi, W. L., Dai, G., & Zhang, L. (2004). Face recognition based on independent Gabor features and support vector machine. In *Fifth World Congress on Intelligent Control and Automation* (pp. 4030-4033).

Chua, C. S., Han, F., & Ho, Y. K. (2000). 3D Human Face Recognition Using Point Signature. In *Proc. of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000* (pp. 233-238).

Chung, K. C., Kee, S. C., & Kim, S. R. (1999). Face Recognition Using Principal Component Analysis of Gabor Filter Responses. In *International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems* (pp. 53-57). Corfu, Greece.

Cristianini Nello & Shawe-Taylor John (2000). *An introduction to Support Vector Machinesand Other Kernel-based Learning Methods*. Cambridge University Press.

Daubechies, I. (1990). The Wavelet Transform, Time-Frequency Localization and Signal Analysis. *IEEE Transactions on Information Theory, 36,* 961-1005.

Daugman, J. G. (1985). Uncertainty Relation for Resolution in Space, Spatial-Frequency, and Orientation Optimized by Two-Dimensional Visual Cortical Filters. *Journal of the Optical Society of America A-Optics Image Science and Vision, 2,* 1160-1169.

Draper, B. A., Baek, K., Bartlett, M. S., & Beveridge, J. R. (2003). Recognizing faces with PCA and ICA. *Computer Vision and Image Understanding, 91,* 115-137.

Duc, B., Fischer, S., & Bigun, J. (1999). Face authentication with Gabor information on deformable graphs. *IEEE Transactions on Image Processing, 8,* 504-516.

Er, M. J., Wu, S. Q., Lu, J. W., & Toh, H. L. (2002). Face recognition with radial basis function (RBF) neural networks. *IEEE Transactions on Neural Networks, 13,* 697-710.

Escobar, M. J. & Ruiz-del-Solar, J. (2002). Biologically-based face recognition using Gabor filters and log-polar images. In *Int. Joint Conf. on Neural Networks - IJCNN* (pp. 1143-1147). Honolulu, USA.

Fan, W., Wang, Y. H., Liu, W., & Tan, T. N. (2004). Combining null space-based Gabor features for face recognition. *Proceedings of the 17Th International Conference on Pattern Recognition, Vol 1,* 330-333.

Fasel, I. R., Barlett, M. S., & Movellan, J. R. (2002). A comparison of Gabor filter methods for automatic detection of facial landmarks. In *Proceedings of the 5th IEEE Int. Conf. on Automatic Face and Gesture Recognition,* (pp. 231-235).

Fisher, R. A. (1936). The use of multiple measures in taxonomic problems. *Ann.Eugenics, 7,* 179-188.

Fleming, M. & Conttrell, G. (1990). Categorization of faces using unsupervised feature extraction. In *IEEE International Joint Conference on Neural Networks* (pp. 65-70).

Freund, Y. & Schapire, R. (1999). A short introduction to boosting. *Journal of Japanese Society for Artifical Intelligence, 14,* 771-780.

Fukunnaga, K. (1991). *Introduction to statistical pattern recognition*. (2 ed.) Academic Press.

Gabor, D. (1946). Theory of communications. *Journal of Institution of Electrical Engineers, 93,* 429-457.

Gokberk, B., Irfanoglu, M. O., Akarun, L., & Alpaydin, E. (2003). Optimal Gabor Kernel Selection for Face Recognition. In *Proceedings of the IEEE International Conference on Image Processing* (pp. 677-680). Barcelona, Spain.

Granham, D. B. & Allison, N. M. (1998). Characterizing Virtual Eigensignatures for General Purpose Face Recognition. In H.Wechsler, P. J. Phillips, V. Bruce, F. Fogelman-Soulie, & T. Huang (Eds.), *Face Recognition: From Theory to Applications* (pp. 446-456).

Granlund, G. H. (1978). Search of A General Picture Processing Operator. *Computer Graphics and Image Processing, 8,* 155-173.

Guo, G. D., Li, S. Z., & Chan, K. L. (2001). Support vector machines for face recognition. *Image and Vision Computing, 19,* 631-638.

Gupta, H. & Agrawal, A. K. (2002). An experimental evaluation of linear and kernel-based methods for face recognition. In *Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision, (WACV 2002)* (pp. 13-18).

Hallinan, P. L., Gordon, G. G., Yuille, A. L., Gibilin, P., & Mumford, D. (1999). *Two- and Three- Dimensional Patterns of the Face*. A K Peters.

Hamamoto, Y., Uchimura, S., Watanabe, M., Yasuda, T., Mitani, Y., & Tomita, S. (1998). A Gabor filter-based method for recognizing handwritten numerals. *Pattern Recognition, 31,* 395-400.

Heisele, B., Ho, P., & Poggio, T. (2001). Face recognition with support vector machines: global versus component-based approach. In *International Conference on Computer Vision (ICCV '01)* (pp. 688-694). Vancouver, Canada.

Hesher, C. S. A. & Erlebacher, G. (2002). PCA of Range Images for Facial Recognition. In *2002 International Multiconference in Computer Science*.

Hjelmas, E. (2000). Feature-based face recognition. In *Proceedings of NOBIM (Norwegian Image Processing and Pattern Recognition Conference)*. Available on line: www.hig.no/~erikh/papers/nobim2000.pdf.

Huang, J. & Heisele, B. (2003). component-based face recognition with morphable models. In *Proc. of the 4th International Conference on Audio- and Video-Based Biometric Person Authetication* (pp. 27-34). Surrey, UK.

Huang, J., Shao, X., & Wechsler, H. (1998). H. Pose discrimination and eye detection using support vector machines. In *Proceeding of NATO-ASI on Face Recognition: From Theory to Applications* (pp. 528-536).

Intel Corporation (2005). OpenCV library. http://www.intel.com/research/mrl/research/opencv/ [On-line].

Jain, A. K. & Farrokhnia, F. (1991). Unsupervised Texture Segmentation Using Gabor Filters. *Pattern Recognition, 24,* 1167-1186.

Jesorsky, O., Kirchberg, K., & Frichholz, R. (2001). Robust Face Detection Using the Hausdorff Distance. In J.Bigun & F. Smeraldi (Eds.), *Proc. of 3rd Audio and Video based Person Authentication* (pp. 90-95).

Jiao, F., Gao, W., Chen, X., Cui, G., & Shan, S. (2002). A face recognition method based on local feature analysis. In *Proc. of the 5th Asian Conference on Computer Vision* (pp. 188-192).

Joachims, T. (2004). SVM-Light package. http://svmlight.joachims.org/. [On-line].

Jones, J. P. & Palmer, L. A. (1987). An Evaluation of the Two-Dimensional Gabor Filter Model of Simple Receptive-Fields in Cat Striate Cortex. *Journal of Neurophysiology, 58,* 1233-1258.

Jonsson, K., Kittler, J., Li, Y. P., & Matas, J. (2002). Support vector machines for face authentication. *Image and Vision Computing, 20,* 369-375.

Kalocsai, P., Zhao, W., & Elagin, E. (1998). Face similarity space as perceived by humans and artifical systems. In *Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 177-180). Japan.

Kepenekci, B. (2001). *Face recognition using Gabor wavelet transform.* M.S. thesis, Dept. of electrical and electronic engineering, The Middle Ease Technical University.

Kepenekci, B., Tek, F. B., & Akar, G. B. (2002). Occluded face recognition based on gabor wavelets. In *Proc. of the IEEE International Conference on Image Processing* (pp. 293-296).

Kim, K. I., Jung, K., & Kim, H. J. (2002). Face recognition using kernel principal component analysis. *IEEE Signal Processing Letters, 9,* 40-42.

Kirby, M. & Sirovich, L. (1990). Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12,* 103-108.

Kruger, V. & Sommer, G. (2000). *Gabor wavelet networks for object representation* Institute of Computer Science, University of Kiel.

Kruger, V. & Sommer, G. (2002a). Gabor wavelet networks for efficient head pose estimation. *Image and Vision Computing, 20,* 665-672.

Kruger, V. & Sommer, G. (2002b). Wavelet networks for face processing. *Journal of the Optical Society of America A-Optics Image Science and Vision, 19,* 1112-1119.

Kyrki, V., Kamarainen, J. K., & Kalviainen, H. (2004). Simple Gabor feature space for invariant object recognition. *Pattern Recognition Letters, 25,* 311-318.

Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., Van der malsburg, C., Wurtz, R. P. et al. (1993). Distortion invariant object recognition in the Dynamic Link Architecture. *IEEE Transactions on Computers, 42,* 300-311.

Lanitis, A., Taylor, C. J., & Cootes, T. F. (1997). Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19,* 743-756.

Lee, C. J. & Wang, S. D. (1999). Fingerprint feature extraction using Gabor filters. *Electronics Letters, 35,* 288-290.

Lee, M. W. & Ranganath, S. (2003). Pose-invariant face recognition using a 3D deformable model. *Pattern Recognition, 36,* 1835-1846.

Li, S. Z. & Zhang, Z. Q. (2004). FloatBoost learning and statistical face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 26,* 1112-1123.

Liao, R. & Li, S. (2000). Face recognition based on multiple facial features. In *Proc. of the 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition* (pp. 239-244).

Lienhart, R. & Maydt, J. (2002). An extended set of Haar-like features for rapid object detection. In *Proc. IEEE Conference on Image Processing* (pp. 900-903).

Liu, C. J. (2004a). Gabor-based kernel PCA with fractional power polynomial models for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 26,* 572-581.

Liu, C. J. & Wechsler, H. (2002). Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing, 11,* 467-476.

Liu, C. J. & Wechsler, H. (2003). Independent component analysis of Gabor feature's for face recognition. *IEEE Transactions on Neural Networks, 14,* 919-928.

Liu, D. H., Lam, K. M., & Shen, L. S. (2004). Optimal sampling of Gabor features for face recognition. *Pattern Recognition Letters, 25,* 267-276.

Liu, Y. H. (2004b). *Face recognition and face detection based on wavelets and neural networks.* Ph.D. Thesis University of Nottingham.

Mallet, S. & Zhong, S. (1992). Wavelet transform maxima and multiscale edges. *Wavelets and Their Applications*.

Menser, B. & Muller, F. (1999). Face detection in color images using principal components analysis. In *Proc.of Seventh International Conference on Image Processing and its Applications* (pp. 620-624).

Messer, K., Kittler, J., Sadeghi, M., Hamouz, M., Kostin, A., Cardinaux, F. et al. (2004). Face authentication test on the BANCA database. In *Proc. of International Conference on Pattern Recognition* (pp. 523-532). Cambridge, UK.

Messer, K., Matas, J., Kittler, J., Luettin, J., & Maitre, G. (1999). Xm2vtsdb: The extended m2vts database. In *Second International Conference on Audio and Video-based Biometric Person Authentication* (pp. 72-77).

Michael, J. & Viola, P. (2003). *Face recognition using boosted local features* (Rep. No. TR2003-025). MItsubishi Electric Research Laboratories.

Moghaddam, B. & Pentland, A. (1994). Face Recognition Using View-Based and Modular Eigenspaces. *Automatic Systems for the Identification and Inspection of Humans, 2277,* 12-31.

Moghaddam, B. & Yang, M. (2000). Gender classification with support vector machines. In *Proceedings. Fourth IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 306-311).

Morimoto, C. H., Koons, D., Amir, A., & Flickner, M. (2000). Pupil detection and tracking using multiple light sources. *Image and Vision Computing, 18,* 331-335.

Mu, X. Y. & Hassoun, M. H. (2003). Combining Gabor features: Summing vs.voting in human face recognition. *2003 IEEE International Conference on Systems, Man and Cybernetics, Vols 1-5, Conference Proceedings,* 737-743.

Muller, K. R., Mika, S., Ratsch, G., Tsuda, K., & Scholkopf, B. (2001). An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks, 12,* 181-201.

Nefian, A. & Hayes, M. (1999). An embedded hmm-based approach for face detection and recognition. In *Proc.of IEEE Int. Conf. On Acoustics, Speech and Signal Processing* (pp. 3553-3556).

Okajima, K. (1998). Two-dimensional Gabor-type receptive field as derived by mutual information maximization. *Neural Networks, 11,* 441-447.

Osuna, E., Freund, R., & Girosit, F. (1997). Training support vector machines: an application to face detection. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 130-136).

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

Phillips, P. J. (1999). Support vector machines applied to face recognition. In *Proceedings of the 1998 conference on Advances in neural information processing systems II* (pp. 803-809). MIT press.

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

Qian, s. & Chen, D. (1996). *Joint Time-Frequency Analysis: Method and Applications*. Prentice Hall PTR.

Reisfeld, D., Wolfson, H., & Yeshurun, Y. (1995). Context-Free Attentional Operators - the Generalized Symmetry Transform. *International Journal of Computer Vision, 14,* 119-130.

Rizon, M. & Kawaguchi, T. (2000). Automatic eye detection using intensity and edge information. In *Proceedings of TENCON* (pp. 24-27).

Romdhani, S., Blanz, V., & Vetter, T. (2002). Face Identification by fitting a 3D Morphable Model using Linear Shape and Texture Error Functions. In *European Conference on Computer Vision* (pp. 3-19).

Ronald, N. B. (1978). *The Fourier Transform and its Applications*. McGraw-Hill, Inc.

Rong, D., Su, G. D., & Lin, X. G. (2002). Face recognition algorithm using local and global information. *Electronics Letters, 38,* 363-364.

Rowley, H. A., Baluja, S., & Kanade, T. (1998). Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 23-38.

Samaria, F. & Young, S. (1994). Hmm-Based Architecture for Face Identification. *Image and Vision Computing, 12,* 537-543.

Schapire, R. E. & Singer, Y. (1999). Improved boosting algorithms using confidence-rated predictions. *Machine Learning, 37,* 297-336.

Scholkopf, B., Mika, S., Burges, C. J. C., Knirsch, P., Muller, K. R., Ratsch, G. et al. (1999). Input space versus feature space in kernel-based methods. *IEEE Transactions on Neural Networks, 10,* 1000-1017.

Scholkopf, B., Smola, A., & Muller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation, 10,* 1299-1319.

Scholkopf, B., Sung, K. K., Burges, C. J. C., Girosi, F., Niyogi, P., Poggio, T. et al. (1997). Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Transactions on Signal Processing, 45,* 2758-2765.

Shan, S. G., Gao, W., Chang, Y. Z., Cao, B., & Yang, P. (2004). Review the strength of Gabor features for face recognition from the angle of its robustness to mis-alignment. *Proceedings of the 17Th International Conference on Pattern Recognition, Vol 1,* 338-341.

Shen, L. & Bai, L. (2004a). AdaBoost Gabor Feature Selection for Classification. In *Proc. of Image and Vision Computing NewZealand* (pp. 77-83).

Shen, L. & Bai, L. (2004b). Face recognition based on Gabor features using kernel methods. In *Proc. of the 6th IEEE Conference on Face and Gesture Recognition* (pp. 170-175). Korea.

Shen, L. & Bai, L. (2004c). Gabor wavelets and Kernel direct discriminant analysis for face recognition. In *Proc. of the 17th International Conference on Pattern Recognition* (pp. 284-287). Cambridge, UK.

Strang, G. & Nguyen, T. (1996). *Wavelets and Filter Banks*. Wellesley-Cambridge Press.

Swets, D. L. & Weng, J. J. (1996). Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18,* 831-836.

Tanaka, H. T., Ikeda, M., & Chiaki, H. (1998). Curvature-based face surface recognition using spherical correlation principal directions for curved object recognition. In *Proc. 3rd International Conference on Automated Face and Gesture Recognition* (pp. 372-377).

The AT&T Lab Cambridge (2002). The ORL Database. http://www.uk.research.att.com:pub/data/att_faces.tar.Z [On-line].

Turk, M. & Pentland, A. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience, 3,* 71-86.

Viola, P. & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition* (pp. 511-518). Kauai, Hawaii.

Wang, X. G. & Tang, X. O. (2003). Bayesian face recognition using Gabor features. In *Proceedings of the 2003 ACM SIGMM workshop on Biometrics methods and applications* (pp. 70-73).

Wang, X. L. & Qi, H. R. (2002). Face Recognition Using Optimal Non-orthogonal Wavelet Basis Evaluated by Information Complexity. In *Proc. of the 16th Internatinal Conference on Pattern Recognition* (pp. 164-167).

Wang, Y. J., Chua, C. S., & Ho, Y. K. (2002). Facial feature detection and face recognition from 2D and 3D images. *Pattern Recognition Letters, 23,* 1191-1202.

Weldon, T. P., Higgins, W. E., & Dunn, D. F. (1996). Efficient Gabor filter design for texture segmentation. *Pattern Recognition, 29,* 2005-2015.

Wiskott, L. (1999). The role of topographical constraints in face recognition. *Pattern Recognition Letters, 20,* 89-96.

Wiskott, L., Fellous, J. M., Kruger, N., & von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19,* 775-779.

Wu, H. Y., Yoshida Y., & Shioyama, T. (2002). Optimal Gabor filters for high speed face identification. In *Proc. of Int. Conf. on Pattern Recognition* (pp. 107-110).

Yang, J., Frangi, A. F., & Yang, J. Y. (2004). A new kernel Fisher discriminant algorithm with application to face recognition. *Neurocomputing, 56,* 415-421.

Yang, M. (2002). Kernel Eigenfaces vs. Kernel Fisherfaces: Face recognition using Kernel methods. In *Proc. of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 205-211). Washington, D.C..

Yuille, A. L., Hallinan, P. W., & Cohen, D. S. (1992). Feature-Extraction from Faces Using Deformable Templates. *International Journal of Computer Vision, 8,* 99-111.

Zhang, H. H., Zhang, B. L., Huang, W. M., & Tian, Q. (2005). Gabor wavelet associative memory for face recognition. *IEEE Transactions on Neural Networks, 16,* 275-278.

Zhang, J., Yan, Y., & Lades, M. (1997). Face recognition: eigenface, elastic matching, and neural nets. *Proceedings of the IEEE, 85,* 1423-1435.

Zhang, Q. G. & Benveniste, A. (1992). Wavelet Networks. *IEEE Transactions on Neural Networks, 3,* 889-898.

Zhao, W. & Chellapa, R. (2000). 3D model enhanced face recognition. In *Proc. International Conference on Image Processing* (pp. 50-53).

Zhao, W., Chellapa, R., Rosenfield, A., & Phillips, P. J. (2000). *Face recognition: A literature survey* (Rep. No. *CVL Technical Report*,). University of Maryland.

Zhao, W., Krishnaswamy, A., Chellapa, R., Swets, D. L., & Weng, J. J. (1998). Discriminant Analysis of Principal Components for Face Recognition. In H.Wechsler, P. J. Phillips, V. Bruce, F. F. Soulie, & Y. P. Huang (Eds.), *Face Recognition: From Theory to Applications* (CAR-TR-914 ed., pp. 73-85). Springer-Verlag.

Zhu, J., Vai, M. I., & Mak, P. U. (2004). Gabor Wavelets Transform and Extended Nearest Feature Space Classifier for Face Recognition. In *Proceedings of the Third International Conference on Image and Graphics* (pp. 246-249).