FACE RECOGNITION USING GABOR WAVELET TRANSFORM

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL SCIENCES

OF

THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

BURCU KEPENEKCI

IN PARTIAL FULLFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

SEPTEMBER 2001

ABSTRACT

FACE RECOGNITION USING GABOR WAVELET TRANSFORM

Kepenekci, Burcu
M.S, Department of Electrical and Electronics Engineering
Supervisor: A. AydınAlatan
Co-Supervisor: GözdeBozdağı Akar

September 2001, 118 pages

Face recognition is emerging as an active research area with numerous commercial and law enforcement applications. Although existing methods performs well under certain conditions, the illumination changes, out of plane rotations and occlusions are still remain as challenging problems. The proposed algorithm deals with two of these problems, namely occlusion and illumination changes. In our method, Gabor wavelet transform is used for facial feature vector construction due to its powerful representation of the behavior of receptive fields in human visual system (HVS). The method is based on selecting peaks (high-energized points) of the Gabor wavelet responses as feature points. Compared to predefined graph nodes of elastic graph matching, our approach has better representative capability for Gabor wavelets. The feature points are automatically extracted using the local characteristics of each individual face in order to decrease the effect of occluded features. Since there is no training as in neural network approaches, a single frontal face for each individual is enough as a reference. The experimental results with standard image libraries, show that the proposed method performs better compared to the graph matching and eigenface based methods.

Keywords: Automatic face recognition, Gabor wavelet transform, human face perception.

ÖZ

GABOR DALGACIKLARINI KULLANARAK YÜZ TANIMA

Kepenekci, Burcu

YüksekLisans, Elektrik Elektronik Mühendisliği Bölümü

Tez Yöneticisi: A. AydınAlatan

Yardımcı Tez Yöneticisi: Gözde Bozdağı Akar

Eylül 2001, 118 sayfa

Yüz tanıma günümüzde hem ticari hem de hukuksal alanlarda artan sayıda uygulaması olan bir problemdir. Varolan yüz tanıma metodları kontrollü ortamda başarılı sonuçlar verse de örtme, yönlenme, ve aydınlatma değişimleri hala yüz tanımada çözülememiş üç problemdir. Önerilen metod ile bu üç problemden aydınlanma değişimleri ve örtme etkisi ele alınmıştır. Bu çalışmada hem yüze ait öznitelik noktaları hem de vektörleri Gabor dalgacık dönüşümü

kullanılarak bulunmuştur. Gabor dalgacık dönüşümü, insan görme sistemindeki duyumsal bölgelerin davranışını modellemesinden dolayı kullanılmıştır. Önerilen metod, daha önceden tanımlanmış çizge düğümleri yerine, Gabor dalgacık tepkeleri tepelerinin (yüksek enerjili noktalarının) öznitelik noktaları olarak seçilmesine dayanmaktadır. Böylece Gabor dalgacıklarının en verimli şekilde kullanılması sağlanmıştır. Öznitelik noktaları otomatik olarak her yüzün farklı yerel özellikleri kullanılarak bulunmakta, bunun sonucu olarak örtük özniteliklerin etkisi de azaltılmaktadır. Sinir ağları yaklaşımlarında olduğu gibi öğrenme safhası olmaması nedeniyle tanıma için her kişinin sadece bir ön yüz görüntüsü yeterli olmaktadır. Yapılan deneylerin sonucunda önerilen metodun varolan çizge eşleme ve özyüzler yöntemleriyle karşılaştırıldığında daha başarılı sonuçlar verdiği gözlenmiştir.

Anahtar Kelimeler: Yüz tanıma, Gabor dalgacık dönüşümü, insan yüz algısı, öznitelik bulma, öznitelik eşleme, örüntü tanıma.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Assoc. Prof. Dr. Gözde Bozdağı Akar and Assist. Prof. Dr. A. Aydın Alatan for their guidance, suggestions and insight throughout this research. I also thank to F. Boray Tek for his support, useful discussions, and great friendship. To my family, I offer sincere thanks for their unshakable faith in me. Special thanks to friends in Signal Processing and Remote Sensing Laboratory for their reassurance during thesis writing. I would like to acknowledge Dr. Uğur Murat Leloğlu for his comments. Most importantly, I express my appreciation to Assoc. Prof. Dr. Mustafa Karaman for his guidance during my undergraduate studies and his encouraging me to research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Machine recognition of faces is emerging as an active research area spanning several disciplines such as image processing, pattern recognition, computer vision and neural networks. Face recognition technology has numerous commercial and law enforcement applications. These applications range from static matching of controlled format photographs such as passports, credit cards, photo ID's, driver's licenses, and mug shots to real time matching of surveillance video images [82].

Humans seem to recognize faces in cluttered scenes with relative ease, having the ability to identify distorted images, coarsely quantized images, and faces with occluded details. Machine recognition is much more daunting task. Understanding the human mechanisms employed to recognize faces constitutes a challenge for psychologists and neural scientists. In addition to the cognitive aspects, understanding face recognition is important, since the same underlying mechanisms could be used to build a system for the automatic identification of faces by machine.

A formal method of classifying faces was first proposed by Francis Galton in 1888 [53, 54]. During the 1980’s work on face recognition remained largely dormant. Since the 1990’s, the research interest in face recognition has grown significantly as a result of the following facts:

1. The increase in emphasis on civilian/commercial research projects,
2. The re-emergence of neural network classifiers with emphasis on real time computation and adaptation,
3. The availability of real time hardware,
4. The increasing need for surveillance related applications due to drug trafficking, terrorist activities, etc.

Still most of the access control methods, with all their legitimate applications in an expanding society, have a bothersome drawback. Except for human and voice recognition, these methods require the user to remember a password, to enter a PIN code, to carry a batch, or, in general, require a human action in the course of identification or authentication. In addition, the corresponding means (keys, batches, passwords, PIN codes) are prone to being lost or forgotten, whereas fingerprints and retina scans suffer from low user acceptance. Modern face recognition has reached an identification rate greater than 90% with well-controlled pose and illumination conditions. While this is a high rate for face recognition, it is not comparable to methods using keys, passwords or batches.

1.1. Why face recognition?

Within today's environment of increased importance of security and organization, identification and authentication methods have developed into a key technology in various areas: entrance control in buildings; access control for computers in general or for automatic teller machines in particular; day-to-day affairs like withdrawing money from a bank account or dealing with the post office; or in the prominent field of criminal investigation. Such requirement for reliable personal identification in computerized access control has resulted in an increased interest in biometrics.

Biometric identification is the technique of automatically identifying or verifying an individual by a physical characteristic or personal trait. The term "automatically" means the biometric identification system must identify or verify a human characteristic or trait quickly with little or no intervention from the user. Biometric technology was developed for use in high-level security systems and law enforcement markets. The key element of biometric technology is its ability to identify a human being and enforce security [83].

Biometric characteristics and traits are divided into behavioral or physical categories. Behavioral biometrics encompasses such behaviors as signature and typing rhythms. Physical biometric systems use the eye, finger, hand, voice, and face, for identification.

A biometric-based system was developed by Recognition Systems Inc., Campbell, California, as reported by Sidlauskas [73]. The system was called ID3D Handkey and used the three dimensional shape of a person's hand to distinguish people. The side and top view of a hand positioned in a controlled capture box were used to generate a set of geometric features. Capturing takes less than two seconds and the data could be stored efficiently in a 9-byte feature vector. This system could store up to 20000 different hands.

Another well-known biometric measure is that of fingerprints. Various institutions around the world have carried out research in the field. Fingerprint systems are unobtrusive and relatively cheap to buy. They are used in banks and to control entrance to restricted access areas. Fowler [51] has produced a short summary of the available systems.

Fingerprints are unique to each human being. It has been observed that the iris of the eye, like fingerprints, displays patterns and textures unique to each human and that it remains stable over decades of life as detailed by Siedlarz [74]. Daugman designed a robust pattern recognition method based on 2-D Gabor transforms to classify human irises.

Speech recognition is also offers one of the most natural and less obtrusive biometric measures, where a user is identified through his or her spoken words. AT&T have produced a prototype that stores a person’s voice on a memory card, details of which are described by Mandelbaum [67].

While appropriate for bank transactions and entry into secure areas, such technologies have the disadvantage that they are intrusive both physically and socially. They require the user to position their body relative to the sensor, then pause for a second to declare himself or herself. This pause and declare interaction is unlikely to change because of the fine-grain spatial sensing required. Moreover, since people can not recognize people using this sort of data, these types of identification do not have a place in normal human interactions and social structures.

While the pause and present interaction perception are useful in high security applications, they are exactly the opposite of what is required when building a store that recognizing its best customers, or an information kiosk that remembers you, or a house that knows the people who live there.

A face recognition system would allow user to be identified by simply walking past a surveillance camera. Human beings often recognize one another by unique facial characteristics. One of the newest biometric technologies, automatic facial recognition, is based on this phenomenon. Facial recognition is the most successful form of human surveillance. Facial recognition technology, is being used to improve human efficiency when recognizing faces, is one of the fastest growing fields in the biometric industry. Interest in facial recognition is being fueled by the availability and low cost of video hardware, the ever-increasing number of video cameras being placed in the workspace, and the noninvasive aspect of facial recognition systems.

Although facial recognition is still in the research and development phase, several commercial systems are currently available and research organizations, such as Harvard University and the MIT Media Lab, are working on the development of more accurate and reliable systems.

1.2. Problem Definition

A general statement of the problem can be formulated as follows: given still or video images of a scene, identify one or more persons in the scene using a stored database of faces.

The environment surrounding a face recognition application can cover a wide spectrum from a well-controlled environment to an uncontrolled one. In a controlled environment, frontal and profile photographs are taken, complete with uniform background and identical poses among the participants. These face images are commonly called *mug shots*. Each mug shot can be manually or automatically cropped to extract a normalized subpart called a canonical face image. In a canonical face image, the size and position of the face are normalized approximately to the predefined values and background region is minimal.

General face recognition, a task that is done by humans in daily activities, comes from virtually uncontrolled environment. Systems, which automatically recognize faces from uncontrolled environment, must detect faces in images. Face detection task is to report the location, and typically also the size, of all the faces from a given image and completely a different problem with respect to face recognition.

Face recognition is a difficult problem due to the general similar shape of faces combined with the numerous variations between images of the same face. Recognition of faces from an uncontrolled environment is a very complex task: lighting condition may vary tremendously; facial expressions also vary from time to time; face may appear at different orientations and a face can be partially occluded. Further, depending on the application, handling facial features over time (aging) may also be required.

Although existing methods performs well under constrained conditions, the problems with the illumination changes, out of plane rotations and occlusions are still remains unsolved. The proposed algorithm, deals with two of these three important problems, namely occlusion and illumination changes.

Since the techniques used in the best face recognition systems may depend on the application of the system, one can identify at least two broad categories of face recognition systems [19]:

1. Finding a person within large database of faces (e.g. in a police database).

   (Often only one image is available per person. It is usually not necessary for recognition to be done in real time.)

2. Identifying particular people in real time (e.g. location tracking system).

   (Multiple images per person are often available for training and real time recognition is required.)

In this thesis, we primarily interested in the first case. Detection of face is assumed to be done beforehand. We aim to provide the correct label (e.g. name label) associated with that face from all the individuals in its database in case of occlusions and illumination changes. Database of faces that are stored in a system is called *gallery*. In gallery, there exists only one frontal view of each individual. We do not consider cases with high degrees of rotation, i.e. we assume that a minimal preprocessing stage is available if required.

1.3. Organization of the Thesis

Over the past 20 years extensive research has been conducted by psychophysicists, neuroscientists and engineers on various aspects of face recognition by human and machines. In chapter 2, we summarize the literature on both human and machine recognition of faces.

Chapter 3 introduces the proposed approach based on Gabor wavelet representation of face images. The algorithm is explained explicitly.

Performance of our method is examined on four different standard face databases with different characteristics. Simulation results and their comparisons to well-known face recognition methods are presented in Chapter 4.

In chapter 5, concluding remarks are stated. Future works, which may follow this study, are also presented.

CHAPTER 2

PAST RESEARCH ON FACE RECOGNITION

The task of recognizing faces has attracted much attention both from neuroscientists and from computer vision scientists. This chapter reviews some of the well-known approaches from these both fields.

2.1. Human Face Recognition: Perceptual and Cognitive Aspects

The major research issues of interest to neuroscientists include the human capacity for face recognition, the modeling of this capability, and the apparent modularity of face recognition. In this section some findings, reached as the result of experiments about human face recognition system, that are potentially relevant to the design of face recognition systems will be summarized

One of the basic issues that have been argued by several scientists is the existence of a dedicated face processing system [82, 3]. Physiological evidence indicates that the brain possesses specialized ‘face recognition hardware’ in the form of face detector cells in the inferotemporal cortex and regions in the frontal right hemisphere; impairment in these areas leads to a syndrome known as *prosapagnosia*. Interestingly, prosapagnosics, although unable to recognize familiar faces, retain their ability to visually recognize non-face objects. As a result of many studies scientists come up with the decision that face recognition is not like other object recognition [42].

Hence, the question is what features humans use for face recognition. The results of the related studies are very valuable in the algorithm design of some face recognition systems. It is interesting that when all facial features like nose, mouth, eye etc. are contained in an image, but in different order than ordinary, recognition is not possible for human. Explanation of face perception as the result of holistic or feature analysis alone is not possible since both are true. In human both global and local features are used in a hierarchical manner [82]. Local features provide a finer classification system for face recognition. Simulations show that the most difficult faces for humans to recognize are those faces, which are neither attractive nor unattractive [4]. Distinctive faces are more easily recognized than typical ones. Information contained in low frequency bands used in order to make the determination of the sex of the individual, while the higher frequency components are used in recognition. The low frequency components contribute to the global description, while the high frequency components contribute to the finer details required in the identification task [8, 11, 13]. It has also been found that the upper part of the face is more useful for recognition than the lower part [82].

In [42], Bruce explains an experiment that is realized by superimposing the low spatial frequency Margaret Thatcher's face on the high spatial frequency components of Tony Blair's face. Although when viewed close up only Tony Blair was seen, viewed from distance, Blair disappears and Margaret Thatcher becomes visible. This demonstrates that the important information for recognizing familiar faces is contained within a particular range of spatial frequencies.

Another important finding is that human face recognition system is disrupted by changes in lighting direction and also changes of viewpoint. Although some scientists tend to explain human face recognition system based on derivation of 3D models of faces using shape from shading derivatives, it is difficult to understand why face recognition appears so viewpoint dependent [1]. The effects of lighting change on face identification and matching suggest that representations for face recognition are crucially affected by changes in low level image features.

Bruce and Langton found that negation (inverting both hue and luminance values of an image) effects badly the identification of familiar faces [124]. They also observe that negation has no significant effect on identification and matching of surface images that lacked any pigmented and textured features, this led them to attribute the negation effect to the alteration of the brightness information about pigmented areas. A negative image of a dark-haired Caucasian, for example, will appear to be a blonde with dark skin. Kemp et al. [125] showed that the hue values of these pigmented regions do not themselves matters for face identification. Familiar faces presented in ‘hue negated’ versions, with preserved luminance values, were recognized as well as those with original hue values maintained, though there was a decrement in recognition memory for pictures of faces when hue was altered in this way [126]. This suggests that episodic memory for pictures of unfamiliar faces can be sensitive to hue, though the representations of familiar faces seems not to be. This distinction between memory for pictures and faces is important. It is clear that recognition of familiar and unfamiliar faces is not the same for humans. It is likely that unfamiliar faces are processed inorder to recognize a picture where as familiar faces are fed into the face recognition system of human brain. A detailed discussion of recognizing familiar and unfamiliar faces can be found in [41].

Young children typically recognize unfamiliar faces using unrelated cues such as glasses, clothes, hats, and hairstyle. By the age of twelve, these paraphernalia are usually reliably ignored. Curiously, when children as young as five years are asked to recognize familiar faces, they do pretty well in ignoring paraphernalia. Several other interesting studies related to how children perceive inverted faces are summarized in [6, 7].

Humans recognize people from their own race better than people from another race. Humans may encode an ‘average’ face; these averages may be different for different races and recognition may suffer from prejudice and unfamiliarity with the class of faces from another race or gender [82]. The poor identification of other races is not a psychophysical problem but more likely a psychosocial one. One of the interesting results of the studies to quantify the role of gender in face recognition is that in Japanese population, majority of the women's facial features is more heterogeneous than the men's features. It has also been found that white women's faces are slightly more variable than men's, but that the overall variation is small [9, 10].

2.1.1. Discussion

The recognition of familiar faces plays a fundamental role in our social interactions. Humans are able to identify reliably a large number of faces and psychologists are interested in understanding the perceptual and cognitive mechanisms at the base of the face recognition process. Those researches illuminate computer vision scientists' studies.

We can summarize the founding of studies on human face recognition system as follows:

1. The human capacity for face recognition is a dedicated process, not merely an application of the general object recognition process. Thus artificial face recognition systems should also be face specific.
2. Distinctive faces are more easily recognized than typical ones.
3. Both global and local features are used for representing and recognizing faces.
4. Humans recognize people from their own race better than people from another race. Humans may encode an 'average' face.

5. Certain image transformations, such as intensity negation, strange viewpoint changes, and changes in lighting direction can severely disrupt human face recognition.

Using the present technology it is impossible to completely model human recognition system and reach its performance. However, the human brain has its shortcomings in the total number of persons that it can accurately ‘remember’. The benefit of a computer system would be its capacity to handle large datasets of face images.

The observations and findings about human face recognition system will be a good starting point for automatic face recognition methods. As it is mentioned above an automated face recognition system should be face specific. It should effectively use features that discriminate a face from others, and more as in caricatures it preferably amplifies such distinctive characteristics of face [5,13].

Difference between recognition of familiar and unfamiliar faces must also be noticed. First of all we should find out what makes us familiar to a face. Seeing a face in many different conditions (different illuminations, rotations, expressions…etc.) make us familiar to that face, or by just frequently looking at the same face image can we be familiar to that face? Seeing a face in many different conditions is something related to training however the interesting point is that by using only the same 2D information how we can pass from unfamiliarity to familiarity. Methods, which recognize faces from a single view, should pay attention to this familiarity subject.

Some of the early scientists were inspired by watching bird flight and built their vehicles with mobile wings. Although a single underlying principle, the Bernoulli effect, explains both biological and man-made flight, we note that no modern aircraft has flapping wings. Designers of face recognition algorithms and systems should be aware of relevant psychophysics and neurophysiological studies but should be prudent in using only those that are applicable or relevant from a practical/implementation point of view.

2.2. Automatic Face Recognition

Although humans perform face recognition in an effortless manner, underlying computations within the human visual system are of tremendous complexity. The seemingly trivial task of finding and recognizing faces is the result of millions years of evolution and we are far away from fully understanding how the brain performs this task.

Up to date, no complete solution has been proposed that allow the automatic recognition of faces in real images. In this section we will review existing face recognition systems in five categories: early methods, neural networks approaches, statistical approaches, template based approaches, and feature based methods. Finally current state of the art of the face recognition technology will be presented.

2.2.1. Representation, Matching and Statistical Decision

The performance of face recognition depends on the solution of two problems: representation and matching.

At an elementary level, the image of a face is a two dimensional (2-D) array of pixel gray levels as,

$$x=\{x_{i,j}, i,j \in S\}, \tag{2.1}$$

where S is a square lattice. However in some cases it is more convenient to express the face image, x, as one-dimensional (1-D) column vector of concatenated rows of pixels, as

$$\boldsymbol{x}=[\,x_1, x_2,.....,x_n]^T \tag{2.2}$$

Where $n=\|S\|$ is the total number of pixels in the image. Therefore $x\in R^n$, the n dimensional Euclidean space.

For a given representation, two properties are important: discriminating power and efficiency; i.e. how far apart are the faces under the representation and how compact is the representation.

While many previous techniques represent faces in their most elementary forms of (2.1) or (2.2), many others use a feature vector, $F(x)=[\,f_1(x), f_2(x),...., f_m(x)]^T$, where $f_1(.),f_2(.),...,f_m(.)$ are linear or nonlinear functionals. Feature-based representations are usually more efficient since generally m is much smaller than n.

A simple way to achieve good efficiency is to use an alternative orthonormal basis of R^n. Specifically, suppose e_1, e_2,..., e_n are an orthonormal basis. Then X can be expressed as

$$x = \sum_{i=1}^{n} \tilde{x}_i e_i \tag{2.3}$$

where $\tilde{x}_i \underline{\underline{\Delta}} \langle x, e_i \rangle$ (inner product), and x can be equivalently represented by $\tilde{x} = [\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n]^T$. Two examples of orthonormal basis are the natural basis used in (2.2) with $e_i=[0,...,0,1,0,...,0]^T$, where one is i^{th} position, and the Fourier basis $e_i = \left(\frac{1}{n^{1/2}}\right)\left[1, e^{j2\pi\left(\frac{i}{n}\right)}, e^{j2\pi\left(\frac{2i}{n}\right)}, ..., e^{j2\pi\left(\frac{(n-1)i}{n}\right)}\right]^T$. If for a given orthonormal basis $\tilde{x}_i$ are small when $i \geq m$, then the face vector $\tilde{x}$ can be compressed into an m dimensional vector, $\tilde{x} \cong [\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_m]^T$.

It is important to notice that an efficient representation does not necessarily have good discriminating power.

In the matching problem, an incoming face is recognized by identifying it with a prestored face. For example, suppose the input face is x and there are K prestored faces c_k, $k=1,2,...,K$. One possibility is to assign x to c_{k_0} if

$$k_0 = \arg\min_{1 \leq k \leq K} \|x - c_k\| \tag{2.4}$$

where $\|.\|$ represents the Euclidean distance in R_n. If $||c_k||$ is normalized so that $||c_k||=c$ for all k, the minimum distance matching in (2.4) simplifies to correlation matching

$$k_0 = \arg\min_{1\le k\le K}\langle x, c_k\rangle . \tag{2.5}$$

Since distance and inner product are invariant to change of orthonormal basis, minimum distance and correlation matching can be performed using any orthonormal basis and the recognition performance will be the same. To do this, simply replace x and c_k in (2.4) or (2.5) by $\tilde{x}$ and $\tilde{c}_k$. Similarly (2.4) and (2.5) also could be used with feature vectors.

Due to such factors such as viewing angle, illumination, facial expression, distortion, and noise, the face images for a given person can have random variations and therefore are better modeled as a random vector. In this case, maximum likelihood (ML) matching is often used,

$$k_0 = \arg\min_{1\le k\le K}\log\left(p\left(x \mid c_k\right)\right) \tag{2.6}$$

where $p(x|c_k)$ is the density of x conditioning on its being the k^{th} person. The ML criterion minimizes the probability of recognition error when a priori, the incoming face is equally likely to be that of any of the K persons. Furthermore if we assume that variations in face vectors are caused by additive white Gaussian noise (AWGN)

$$x_k=c_k+w_k \tag{2.7}$$

where w_k is a zero-mean AWGN with power σ^2, then the ML matching becomes the minimum distance matching of (2.4).

2.2.2. Early face recognition methods

The initial work in automatic face processing dates back to the end of the 19th century, as reported by Benson and Perrett [39]. In his lecture on personal identification at the Royal Institution on 25 May 1888, Sir Francis Galton [53], an English scientist, explorer and a cousin of Charles Darwin, explained that he had "frequently chafed under the sense of inability to verbally explain hereditary resemblance and types of features". In order to relieve himself from this embarrassment, he took considerable trouble and made many experiments. He described how French prisoners were identified using four primary measures (head length, head breadth, foot length and middle digit length of the foot and hand respectively). Each measure could take one of the three possible values (large, medium, or small), giving a total of 81 possible primary classes. Galton felt it would be advantageous to have an automatic method of classification. For this purpose, he devised an apparatus, which he called a *mechanical selector*, that could be used to compare measurements of face profiles. Galton reported that most of the measures he had tried were fairy efficient.

Early face recognition methods were mostly feature based. Galton's proposed method, and a lot of work to follow, focused on detecting important facial features as eye corners, mouth corners, nose tip, etc. By measuring the relative distances between these facial features a feature vector can be constructed to describe each face. By comparing the feature vector of an unknown face to the feature vectors of known vectors from a database of known faces, the closest match can be determined.

One of the earliest works is reported by Bledsoe [84]. In this system, a human operator located the feature points on the face and entered their positions into the computer. Given a set of feature point distances of an unknown person, nearest neighbor or other classification rules were used for identifying the test face. Since feature extraction is manually done, this system could accommodate wide variations in head rotation, tilt, image quality, and contrast.

In Kanade's work [62], series fiducial points are detected using relatively simple image processing tools (edge maps, signatures etc.) and the Euclidean distances are then used as a feature vector to perform recognition. The face feature points are located in two stages. The coarse-grain stage simplified the succeeding differential operation and feature finding algorithms. Once the eyes, nose and mouth are approximately located, more accurate information is extracted by confining the processing to four smaller groups, scanning at higher resolution, and using 'best beam intensity' for the region. The four regions are the left and right eye, nose, and mouth. The beam intensity is based on the local area histogram obtained in the coarse-gain stage. A set of 16 facial parameters, which are rations of distances, areas, and angles to compensate for the varying size of the pictures, is extracted. To eliminate scale and dimension differences the components of the resulting vector are normalized. A simple distance measure is used to check similarity between two face images.

2.2.3. Statistical approaches to face recognition

2.2.3.1. Karhunen-Loeve Expansion Based Methods

2.2.3.1.1. Eigenfaces

A face image, *I(x,y)*, of size *NxN* is simply a matrix with beach element representing the intensity at that particular pixel. *I(x,y)* may also be considered as a vector of length N^2 or a single point in an N^2dimentional space. So a *128x128* pixel image can be represented as a point in a 16,384 dimensional space. Facial images in general will occupy only a small sub-region of this high dimensional ‘image space’ and thus are not optimally represented in this coordinate system.

As mentioned in section 2.2.1, alternative orthonormal bases are often used to compress face vectors. One such basis is the Karhunen-Loeve (KL).

The ‘Eigenfaces’ method proposed by Turk and Pentland [20], is based on the Karhunen-Loeve expansion and is motivated by the earlier work of Sirovitch and Kirby [63] for efficiently representing picture of faces. Eigenface recognition derives it is name from the German prefix ‘eigen’, meaning ‘own’ or ‘individual’. The Eigenface method of facial recognition is considered the first working facial recognition technology.

The eigenfaces method presented by Turk and Pentland finds the principal components (Karhunen-Loeve expansion) of the face image distribution or the eigenvectors of the covariance matrix of the set of face images. These eigenvectors can be thought as a set of features, which together characterize the variation between face images.

Let a face image $I(x,y)$ be a two dimensional array of intensity values, or a vector of dimension n. Let the training set of images be $I_1, I_2, ..., I_N$. The average face image of the set is defined by $\psi = \frac{1}{N}\sum_{i=1}^{N} I_i$. Each face differs from the average by the vector $\phi_i = I_i - \psi$. This set of very large vectors is subject to principal component analysis which seeks a set of K orthonormal vectors v_k, $k=1,...,K$ and their associated eigenvalues λ_k which best describe the distribution of data.

Vectors v_k and scalars λ_k are the eigenvectors and eigenvalues of the covariance matrix:

$$C = \frac{1}{N}\sum_{i=1}^{N} \phi_i \phi_i^T = AA^T , \qquad (2.9)$$

where the matrix $A=[\phi_1, \phi_2,..., \phi_N]$. Finding the eigenvectors of matrix C_{nxn} is computationally intensive. However, the eigenvectors of C can be determined by first finding the eigenvectors of a much smaller matrix of size NxN and taking a linear combination of the resulting vectors.

$$Cv_k = \lambda_k v_k \qquad (2.10)$$

$$v_k{}^T C v_k = v_k{}^T \lambda_k v_k \qquad (2.11)$$

since eigenvectors, v_k, are orthogonal and normalized $v_k^T v_k=1$.

$$v_k{}^T C v_k = \lambda_k \qquad (2.12)$$

$$\lambda_k = \frac{1}{N} v_k{}^T \sum_{i=1}^{N} \phi_i \phi_i^T v_k \tag{2.13}$$

$$= \frac{1}{N} \sum_{i=1}^{N} v_k{}^T \phi_i \phi_i^T v_k$$

$$= \frac{1}{N} \sum_{i=1}^{N} (v_k \phi_i^T)^T (v_k \phi_i^T)$$

$$= \frac{1}{N} \sum_{i=1}^{N} (v_k \phi_i^T)^2$$

$$= \frac{1}{N} \sum_{i=1}^{N} (v_k I_i^T - mean(v_k I_i^T))^2$$

$$= \frac{1}{N} \sum_{i=1}^{N} \mathrm{var}(v_k I_i^T)$$

Thus eigenvalue k represents the variance of the representative facial image set along the axis described by eigenvector k.

The space spanned by the eigenvectors v_k, $k=1,...,K$ corresponding to the largest K eigenvalues of the covariance matrix C, is called the *face space*. The eigenvectors of matrix C, which are called eigenfaces from a basis set for the face images. A new face image Γ is transformed into its eigenface components (projected onto the face space) by:

$$\omega_k = < vk, (\Gamma - \phi) > = vk^T (\Gamma - \phi) \tag{2.14}$$

For $k=1,...,K$. The projections w_k form the feature vector $\Omega = [\ w_1,\ w_2, ...,\ w_K]$ which describes the contribution of each of each eigenface in representing the input image.

Given a set of face classes E_q and the corresponding feature vectors Ωq, the simplest method for determining which face class provides the best description of an input face image Γ is to find the face class that minimizes the Euclidean distance in the feature space:

$$\xi_q = \left\| \Omega - \Omega_q \right\| \tag{2.15}$$

A face is classified as belonging to class E_q when the minimum ξq is below some threshold θ_ϵ and also

$$E_q = \arg\min_q \{\xi_q\}. \tag{2.16}$$

Otherwise, the face is classified as unknown.

Turk and Pentland [20] test how their algorithm performs in changing conditions, by varying illumination, size and orientation of the faces. They found that their system had the most trouble with faces scaled larger or smaller than the original dataset. To overcome this problem they suggest using a multi-resolution method in which faces are compared to eigenfaces of varying sizes to compute the best match. Also they note that image background can have significant effect on performance, which they minimize by multiplying input images with a 2-D Gaussian to diminish the contribution of the background and highlight the central facial features. System performs face recognition in real-time. Turk and Pentland's paper was very seminal in the field of face recognition and their method is still quite popular due to its ease of implementation.

Murase and Nayar [85] extended the capabilities of the eigenface method to general 3D-object recognition under different illumination and viewing conditions. Given N object images taken under P views and L different illumination conditions, a universal image set is built which contains all the available data. In this way a single ‘parametric space’ describes the object identity as well as the viewing or illumination conditions. The eigenface decomposition of this space was used for feature extraction and classification. However in order to insure discrimination between different objects the number of eigenvectors used in this method was increased compared to the classical Eigenface method.

Later, based on the eigenface decomposition, Pentland et al [86] developed a ‘view based’ eigenspace approach for human face recognition under general viewing conditions. Given *N* individuals under *P* different views, recognition is performed over *P* separate eigenspaces, each capturing the variation of the individuals in a common view. The ‘view based’ approach is essentially an extension of the eigenface technique to multiple sets of eigenvectors, one for each face orientation. In order to deal with multiple views, in the first stage of this approach, the orientation of the test face is determined and the eigenspace which best describes the input image is selected. This is accomplished by calculating the residual description error (distance from feature space: DFFS) for each view space. Once the proper view is determined, the image is projected onto appropriate view space and then recognized. The view based approach is computationally more intensive than the parametric approach because *P* different sets of *V* projections are required (*V* is the number of eigenfaces selected to represent each eigenspace). Naturally, the view-based representation can yield more accurate representation of the underlying geometry.

2.2.3.1.2. Face Recognition using Eigenfaces

There are two main approaches of recognizing faces by using eigenfaces.

Appearance model:

1- A database of face images is collected

2- A set of eigenfaces is generated by performing principal component analysis (PCA) on the face images. Approximately, 100 eigenvectors are enough to code a large database of faces.

3- Each face image is represented as a linear combination of the eigenfaces.

4- A given test image is approximated by a combination of eigenfaces. A distance measure is used to compare the similarity between two images.

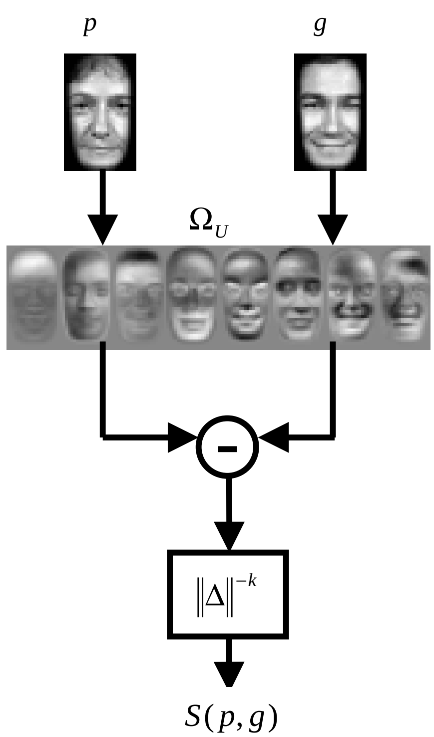


Figure 2.1: Appearance model

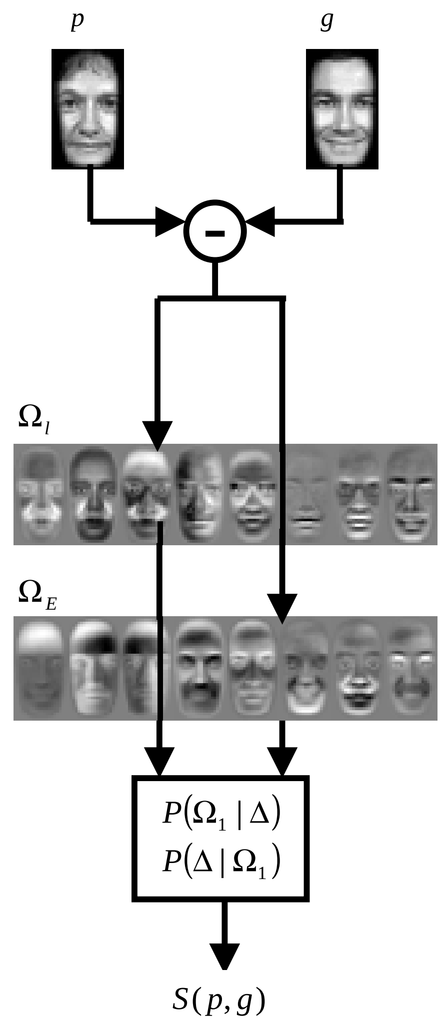


Figure 2.2: Discriminative model

Discriminative model:

1- Two datasets Ω_I and Ω_E are obtained by computing intrapersonal differences (by matching two views of each individual in the dataset) and the other by computing extrapersonal differences (by matching different individuals in the dataset), respectively.

2- Two datasets of eigenfaces are generated by performing PCA on each class.

3- Similarity score between two images is derived by calculating $S=P(\Omega_I|\Delta)$, where Δ is the difference between a pair of images. Two images are determined to be the same individual, if $S>0.5$.

Although the recognition performance is lower than the correlation method, the substantial reduction in computational complexity of the eigenface method makes this method very attractive. The recognition rates increase with the number of principal components (eigenfaces) used and in the limit, as more principal components are used, performance approaches that of correlation. In [20], and [86], authors reported that the performances level off at about 45 principal components.

It has been shown that removing first three principal components results in better recognition performances (the authors reported an error rate of %20 when using the eigenface method with 30 principal components on a database strongly affected by illumination variations and only %10 error rate after removing the first three components). The recognition rates in this case were better than the recognition rates obtained using the correlation method. This was argued based on the fact that first components are more influenced by variations of lighting conditions.

2.2.3.1.3. Eigenfeatures

Pentland et al. [86] discussed the use of facial features for face recognition. This can be either a modular or a layered representation of the face, where a coarse (low-resolution) description of the whole head is augmented by additional (high-resolution) details in terms of salient facial features. The eigenface technique was extended to detect facial features. For each of the facial features, a feature space is built by selecting the most significant eigenfeatures (eigenvectors corresponding to the largest eigenvalues of the features correlation matrix).

After the facial features in a test image were extracted, a score of similarity between the detected features and the features corresponding to the model images is computed. A simple approach for recognition is to compute a cumulative score in terms of equal contribution by each of the facial feature scores. More elaborate weighting schemes can also be used for classification. Once the cumulative score is determined, a new face is classified such that this score is maximized.

The performance of eigenfeatures method is close to that of eigenfaces, however a combined representation of eigenfaces and eigenfeatures shows higher recognition rates.

2.2.3.1.4. The Karhunen-Loeve Transform of the Fourier Spectrum

Akamatsu et. al. [87], illustrated the effectiveness of Karhunen-Loeve Transform of Fourier Spectrum in the Affine Transformed Target Image (KL-FSAT) for face recognition. First, the original images were standardized with respect to position, size, and orientation using an affine transform so that three reference points satisfy a specific spatial arrangement. The position of these points is related to the position of some significant facial features. The eigenface method is applied discussed in the section 2.2.3.1.1. to the magnitude of the Fourier spectrum of the standardized images (KL-FSAT). Due to the shift invariance property of the magnitude of the Fourier spectrum, the KL-FSAT performed better than classical eigenfaces method under variations in head orientation and shifting. However the computational complexity of KL-FSAT method is significantly greater than the eigenface method due to the computation of the Fourier spectrum.

2.2.3.2. Linear Discriminant Methods- Fisherfaces

In [88], [89], the authors proposed a new method for reducing the dimensionality of the feature space by using Fisher's Linear Discriminant (FLD) [90]. The FLD uses the class membership information and develops a set of feature vectors in which variations of different faces are emphasized while different instances of faces due to illumination conditions, facial expression and orientations are de-emphasized.

2.2.3.2.1. Fisher's Linear Discriminant

Given c classes with a priori probabilities P_i, let N_i be the number of samples of class i, $i=1,...,c$. Then the following positive semi-definite scatter matrices are defined as:

$$S_B = \sum_{i=1}^{c} P_i (\mu_i - \mu)(\mu_i - \mu)^T \qquad (2.17)$$

$$S_w = \sum_{i=1}^{c} P_i \sum_{j=1}^{N_i} (x_j^i - \mu_i)(x_j^i - \mu_i)^T \qquad (2.18)$$

Where x_j^i denotes the j[th] n-dimensional sample vector belonging to class i, μ_i is the mean of class i:

$$\mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_j^i, \qquad (2.19)$$

and μ is the overall mean of sample vectors:

$$\mu = \frac{1}{\sum_{i=1}^{c} N_i} \sum_{i=1}^{c} \sum_{j=1}^{N_i} x_j^i, \qquad (2.20)$$

S_w is the within –classscatter matrix and represents the average scatter of sample vector of class i; S_B is the between-class scatter matrix and represents the scatter of the mean μ_i of class i around the overall mean vector μ. If S_w is non singular, the Linear Discriminant Analysis (LDA) selects a matrix $V_{opt} \in R^{nxk}$ with orthonormal columns which maximizes the ratio of the determinant of the between class scatter matrix of the projected samples,

$$V_{opt} = \arg\max_v \left(\frac{|V^T S_B V|}{|V^T S_w V|} \right) = [v_1, v_2, ..., v_k], \qquad (2.21)$$

Where $\{v_i|i=1,\ldots,k\}$ is the set of generalized eigenvectors of S_B and S_w corresponding to the set of decreasing eigenvalues $\{\lambda_i \,|i=1,\ldots,k\}$, i.e.

$$S_B v_i = \lambda_i S_w v_i. \tag{2.22}$$

As shown in [91], the upper bound of k is c-1. The matrix V_{opt} describes the Optimal Linear Discriminant Transform or the Foley-Sammon Transform. While the Karhunen-Loeve Transform performs a rotation on a set of axes along which the projection of sample vectors differ most in the autocorrelation sense, the Linear Discriminant Transform performs a rotation on a set of axes $[v_1, v_2,\ldots, v_k]$ along which the projection of sample vectors show maximum discrimination.

2.2.3.2.2. Face Recognition Using Linear Discriminant Analysis

Let a training set of N face images represents c different subjects. The face image in the training set are two-dimensional arrays of intensity values, represented as vectors of dimension n. Different instances of a person’s face (variations in lighting, pose or facial expressions) are defined to be in the same class and faces of different subjects are defined to be from different classes.

The scatter matrices S_B and Sw are defined in Equations (2.17), (2.18). However the matrix V_{opt} cannot be found directly from Equation (2.21), because in general matrix S_w is singular. This stems from the fact that the rank of S_w is less than N-c, and in general, the number of pixels in each image n is much larger than the number of images in the learning set N. There have been presented many solutions in the literature in order to overcome this problem [92, 93]. In [88], the authors propose a method which is called *Fisherfaces*. The problem of S_w being singular is avoided by projecting the image set onto a lower dimensional space so that the resulting within class scatter is non singular. This is achieved by using Principal Component Analysis (PCA) to reduce the dimension of the feature space to *N-c* and then, applying the standard linear discriminant defined in Equation (2.21) to reduce the dimension to *c-1*. More formally V_{opt} is given by:

$$V_{opt} = V_{fld} V_{pca}, \tag{2.23}$$

Where

$$V_{pca} = \arg\max_{v} \left[V^T C V \right], \tag{2.24}$$

and,

$$V_{fld} = \arg\max_{v} \left(\frac{\left| V^T V_{pca}^T SB V_{pca} V \right|}{\left| V^T V_{pca}^T Sw V_{pca} V \right|} \right), \tag{2.25}$$

Where C is the covariance matrix of the set of training images and is computed from Equation (2.9). The columns of V_{opt} are orthogonal vectors which are called Fisherfaces. Unlike the Eigenfaces, the Fisherfaces do not correspond to face like patterns. All example face images E_q, $q=1,...,Q$ in the example set S are projected on to the vectors corresponding to the columns of the V_{fld} and a set of features is extracted for each example face image. These feature vectors are used directly for classification.

Having extracted a compact and efficient feature set, the recognition task can be performed by using the Euclidean distance in the feature space. However, in [89] as a measure in the feature space, is proposed a weighted mean absolute/square distance with weights obtained based on the reliability of the decision axis.

$$D(\Gamma,\mathrm{E}) = \sum_{v=1}^{K} \frac{(\Gamma_v - \mathrm{E}_v)^2}{\Sigma_{\mathrm{E}\in S}(\Gamma_v - \mathrm{E}_v)^2}\alpha_v\,, \tag{2.26}$$

Therefore, for a given face image *Γ*, the best match E^0 is given by

$$\mathrm{E}^0 = \arg\min_{\mathrm{E}\in S}\{D(\Gamma,\mathrm{E})\}. \tag{2.27}$$

The confidence measure is defined as:

$$Conf(\Gamma,\mathrm{E}^0) = 1 - \left(\frac{D(\Gamma,\mathrm{E}^0)}{D(\Gamma,\mathrm{E}^1)}\right), \tag{2.28}$$

where E^1 is the second best candidate.

In [87], Akamatsu et. al. applied LDA to the Fourier Spectrum of the intensity image. The results reported by the authors showed that LDA in the Fourier domain is significantly more robust to variations in lighting than the LDA applied directly to the intensity images. However the computational complexity of this method is significantly greater than classical Fisherface method due to the computation of the Fourier spectrum.

2.2.3.3. Singular Value Decomposition Methods

2.2.3.3.1 Singular value decomposition

Methods based on the Singular Value Decomposition for face recognition use the general result stated by the following theorem:

Theorem: Let I_{pxq} be a real rectangular matrix $Rank(I)=r$, then there exists two orthonormal matrices U_{pxp}, V_{qxq} and a diagonal matrix Σ_{pxq} and the following formula holds:

$$I = U\Sigma V^T = \sum_{i=1}^{r} \lambda_i v_i v_i^T, \tag{2.29}$$

where

$U=(u_1, u_2, \ldots, u_r, u_{r+1}, \ldots, u_p)$,

$V=(v_1, v_2, \ldots, v_r, v_{r+1}, \ldots, v_q)$,

$\Sigma=diag(\lambda_1, \lambda_2, \ldots, \lambda_r, 0, \ldots, 0)$,

$\lambda_1>\lambda_2>\ldots>\lambda_r>0$, λ_i^2, $i=1,\ldots,r$ are the eigenvalues of II^T and I^TI, u_i, v_j, $i=1,\ldots,p$, $j=1,\ldots,q$ are the eigenvectors corresponding to eigenvalues of II^T and I^TI.

2.2.3.3.2. Face recognition Using Singular Value Decomposition

Let a face image $I(x,y)$ be a two dimensional (mxn) array of intensity values and $[\lambda_1, \lambda_2, \ldots, \lambda_r]$ be its singular value *(SV)* vector. In [93], Zhong revealed the importance of using SVD for human face recognition by proving several important properties of the *SV* vector as: the stability of the *SV* vector to small perturbations caused by stochastic variation in the intensity image, the proportional variance of the *SV* vector to proportional variance of pixels in the intensity image, the invariance of the *SV* feature vector to rotation transform, translation and mirror transform. The above properties of the *SV* vector provide the theoretical basis for using singular values as image features. However, it has been shown that compressing the original *SV* vector into a low dimensional space, by means of various mathematic transforms leads to higher recognition performances. Among various transformations of compressing dimensionality, the Foley-Sammon transform based on Fisher criterion, i.e. optimal discriminant vectors, is the most popular one. Given *N* face images, which present *c* different subjects, the *SV* vectors are extracted from each image. According to Equations (2.17) and (2.18), the scatter matrices S_B and S_w of the *SV* vectors are constructed. It has been shown that it is difficult to obtain the optimal discriminant vectors in the case of small number of samples, i.e. the number of samples is less than the dimensionality of the *SV* vector because the scatter matrix S_w is singular in this case. Many solutions have been proposed to overcome this problem. Hong [93], circumvented the problem by adding a small singular value perturbation to S_w resulting in $S_w(t)$ such that $S_w(t)$ becomes nonsingular. However the perturbation of S_w introduces an arbitrary parameter, and the range to which the authors restricted the perturbation is not appropriate to ensure that the inversion of $S_w(t)$ is numerically stable. Cheng et al [92], solved the problem by rank decomposition of S_w. This is a generalization of Tian's method [94], who substitute S_w^- by the positive pseudo-inverse S_w^+.

After the set of optimal discriminant vectors $\{v_1, v_2, ..., v_k\}$ has been extracted, the feature vectors are obtained by projecting the *SV* vectors onto the space spanned by $\{v_1, v_2, ..., v_k\}$.

When a test image is acquired, its *SV* vector is projected onto the space spanned by $\{v_1, v_2, ..., v_k\}$ and classification is performed in the feature space by measuring the Euclidean distance in this space and assigning the test image to the class of images for which the minimum distance is achieved.

Another method to reduce the feature space of the *SV* feature vectors was described by Cheng et al [95]. The training set used consisted of a small sample of face images of the same person. If I_j^i represents the j^{th} face image of person i, then the average image is given by $\frac{1}{N}\sum_{j=1}^{N} I_j^i$. Eigenvalues and eigenvectors are determined for this average image using SVD. The eigenvalues are tresholded to disregard the values close to zero. Average eigenvectors (called feature vectors) for all the average face images are calculated. A test image is then projected onto the space spanned by the eigenvectors. The Frobenius norm is used as a criterion to determine which person the test image belongs.

2.2.4. Hidden Markov Model Based Methods

Hidden Markov Models (HMM) are a set of statistical models used to characterize the statistical properties of a signal. Rabiner [69][96], provides an extensive and complete tutorial on HMMs. HMM are made of two interrelated processes:

- an underlying, unobservable Markov chain with finite number of states, a state transition probability matrix and an initial state probability distribution.
- A set of probability density functions associated to each state.

The elements of HMM are:

N, the number of states in the model. If *S* is the set of states, then $S=\{ S_1, S_2,..., S_N\}$. The state of the model q_t time t is given by $q_t \in S$, $1 \le t \le T$, where *T* is the length of the observation sequence (number of frames).

M, the number of different observation symbols. If *V* is the set of all possible observation symbols (also called the codebook of the model), then $V=\{ V_1, V_2,..., V_M\}$.

A, the state transition probability matrix; $A=\{a_{ij}\}$ where

$$A_{ij} = P[q_t = S_j \mid q_{t-1} = S_i\,], \quad 1 \le i, j \le N \tag{2.30}$$

$$0 \le a_{ij} \le 1, \quad \sum_{j=1}^{N} a_{ij} = 1, \quad 1 \le i \le N \tag{2.31}$$

B, the observation symbol probability matrix; $B=b_j(k)$ where,

$$B_j(k) = P[Q_t = vk \mid q_t = S_j\,], \quad 1 \le j \le N, \quad 1 \le k \le M \tag{2.32}$$

And Q_t is the observation symbol at time *t*.

π, the initial state distribution; $\pi=\pi_i$ where

$$\pi_i = P[q_1 = S_i\,], \quad 1 \le i \le N \tag{2.33}$$

Using a shorthand notation, a HMM is defined as:

$$\lambda=(A,B,\pi). \tag{2.34}$$

The above characterization corresponds to a discrete HMM, where the observations characterized as discrete symbols chosen from a finite alphabet $V=\{v_1, v_2,...,v_M\}$. In a continuous density HMM, the states are characterized by continuous observation density functions. The most general representation of the model probability density function (pdf) is a finite mixture of the form:

$$b_i(O)=\sum_{k=1}^{M} c_{ik} N(O,\mu_{ik},U_{ik}), \qquad 1 \le i \le N \tag{2.35}$$

where c_{ik} is the mixture coefficient for the k^{th} mixture in state i. Without loss of generality $N(O,\mu_{ik}, U_{ik})$ is assumed to be a Gaussian pdf with mean vector μ_{ik} and covariance matrix U_{ik}.

HMM have been used extensively for speech recognition, where data is naturally one-dimensional (1-D) along time axis. However, the equivalent fully connected two-dimensional HMM would lead to a very high computational problem [97]. Attempts have been made to use multi-model representations that lead to pseudo 2-D HMM [98]. These models are currently used in character recognition [99][100].

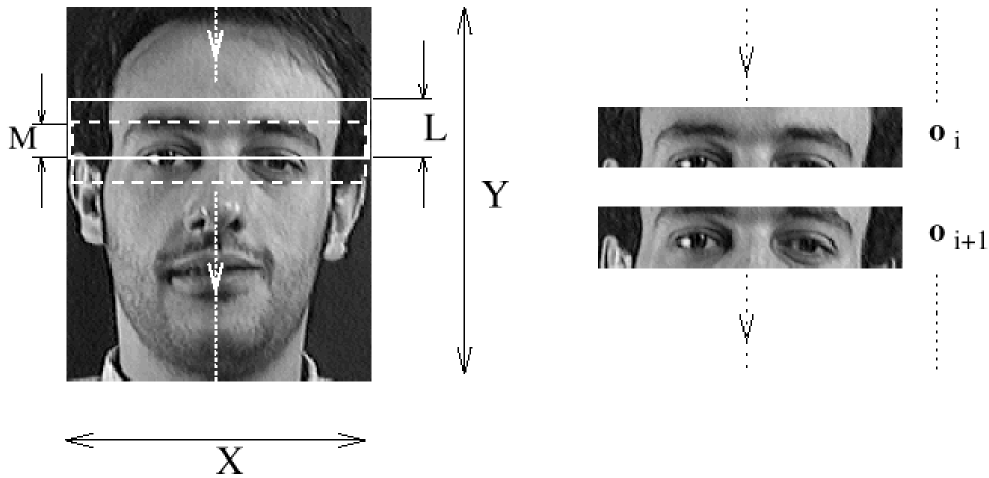


Figure 2.3: Image sampling technique for HMM recognition

In [101], Samaria et al proposed the use of the 1-D continuous HMM for face recognition. Assuming that each face is in an upright, frontal position, features will occur in a predictable order. This ordering suggests the use of a top-bottom model, where only transitions between adjacent states in a top to bottom manner are allowed [102]. The states of the model correspond to the facial features forehead, eyes, nose, mouth and chin [103]. The observation sequence O is generated from an *XxY* image using an *XxL* sampling window with *XxM* pixels overlap (Figure 2.3). Each observation vector is a block of *L* lines. There is an M line overlap between successive observations. The overlapping allows the features to be captured in a manner, which is independent of vertical position, while a disjoint partitioning of the image could result in the truncation of features occurring across block boundaries. In [104], the effect of different sampling parameters has been discussed. With no overlap, if a small height of the sampling window is used, the segmented data do not correspond to significant facial features. However, as the window height increases, there is a higher probability of cutting across the features.

Given c face images for each subject of the training set, the goal of the training stage is to optimize the parameters $\lambda_i=(A,B,\pi)$ to describe 'best', the observations $O=\{ o_1, o_2,...,o_T\}$, in the sense of maximizing $P(O|\lambda)$. The general HMM training scheme is illustrated in Figure 2.4 and is a variant of the K-means iterative procedure for clustering data:

1. The training images are collected for each subject in the database and are sampled to generate the observation sequence.
2. A common prototype (state) model is constructed with the purpose of specifying the number of states in the HMM and the state transitions allowed, A (model initialization).
3. A set of initial parameter values using the training data and the prototype model are computed iteratively. The goal of this stage is to find a good estimate for the observation model probability matrix B. In [96], it has been shown that a good initial estimates of the parameters are essential for rapid and proper convergence (to the global maximum of the likelihood function) of the re-estimation formulas. On the first cycle, the data is uniformly segmented, matched with each model state and the initial model parameters are extracted. On successive cycles, the set of training observation sequences was segmented into states via the Viterbi algorithm [50]. The result of segmenting each of the training sequences is for each of N states, a maximum likelihood estimate of

the set of observations that occur within each state according to the current model.

4. Following the Viterbi segmentation, the model parameters are re-estimated using the Baum-Welch re-estimation procedure. This procedure adjusts the model parameters so as to maximize the probability of observing the training data, given each corresponding model.
5. The resulting model is then compared to the previous model (by computing a distance score that reflects the statistical similarity of the HMMs). If the model distance score exceeds a threshold, then the old model λ is replaced by the new model $\tilde{\lambda}$, and the overall training loop is repeated. If the model distance score falls below the threshold, then model convergence is assumed and the final parameters are saved.

Recognition is carried out by matching the test image against each of the trained models (Figure 2.5). In order to achieve this, the image is converted to an observation sequence and then model likelihoods $P(O_{test}|\lambda_i)$ are computed for each λ_i, $i=1,...,c$. The model with highest likelihood reveals the identity of the unknown face, as

$$V = \arg\max_{1 \le i \le c} \left[P(O_{test} \mid \lambda_i) \right]. \qquad (2.36)$$

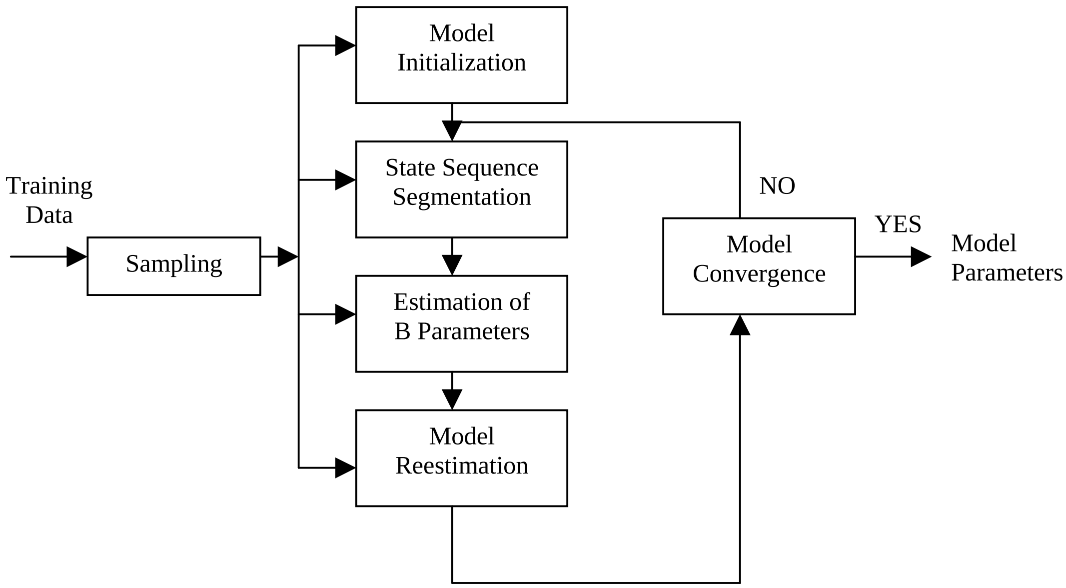


Figure 2.4: HMM training scheme

The HMM based method showed significantly better performances for face recognition compared to the eigenface method. This is due to fact that HMM based method offers a solution to facial features detection as well as face recognition.

However the 1-D continuous HMM are computationally more complex than the Eigenface method. A solution in reducing the running time of this method is the use of discrete HMM. Extremely encouraging preliminary results (error rates below %5) were reported in [105] when pseudo 2-D HMM are used. Furthermore, the authors suggested that Fourier representation of the images can lead to better recognition performance as frequency and frequency-space representation can lead to better data separation.

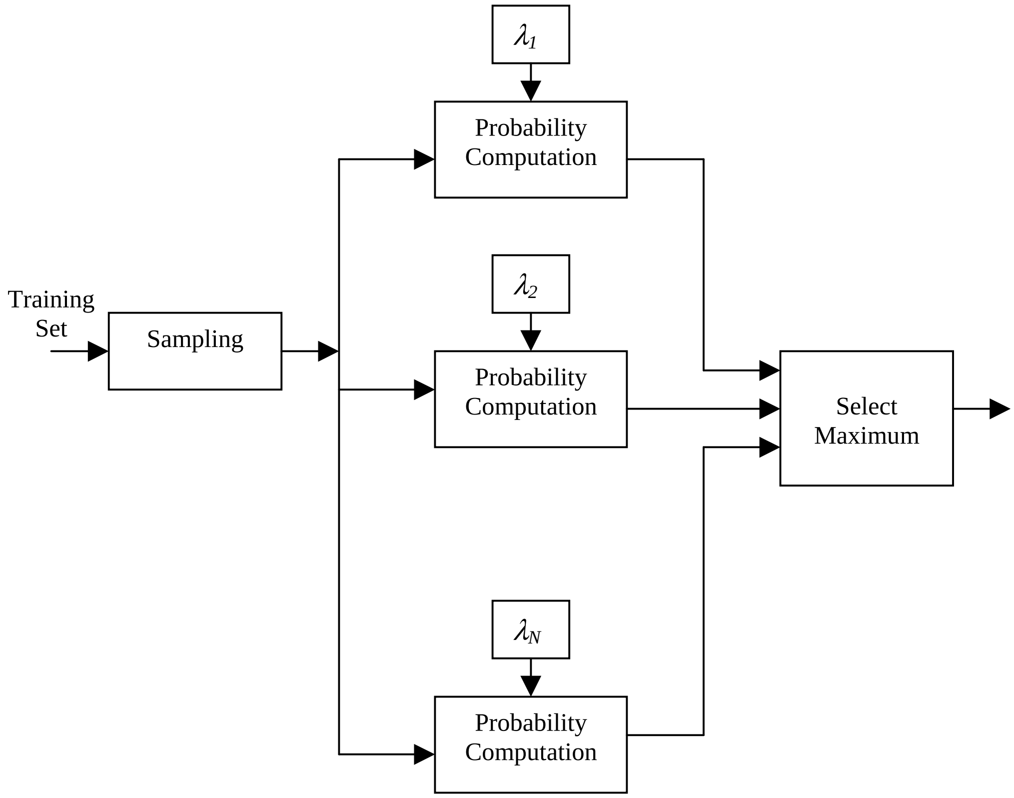


Figure 2.5: HMM recognition scheme

2.2.5. Neural Networks Approach

In principal, the popular back-propagation (BP) neural network [106] can be trained to recognize face images directly. However, a simple network can be very complex and difficult to train. A typical image recognition network requires $N=mxn$ input neurons, one for each of the pixels in an nxm image. For example, if the images are 128x128, the number of inputs of the network would be 16,384. In order to reduce the complexity, Cottrell and Fleming [107] used two BP nets (Figure 2.6). The first net operates in the auto-association mode [108] and extracts features for the second net, which operates in the more common classification mode.

The autoassociation net has n inputs, n outputs and p hidden layer nodes. Usually p is much smaller than n. The network takes a face vector x as an input and is trained to produce an output y that is a 'best approximation' of x. In this way, the hidden layer output h constitutes a compressed version of x, or a feature vector, and can be used as the input to classification net.

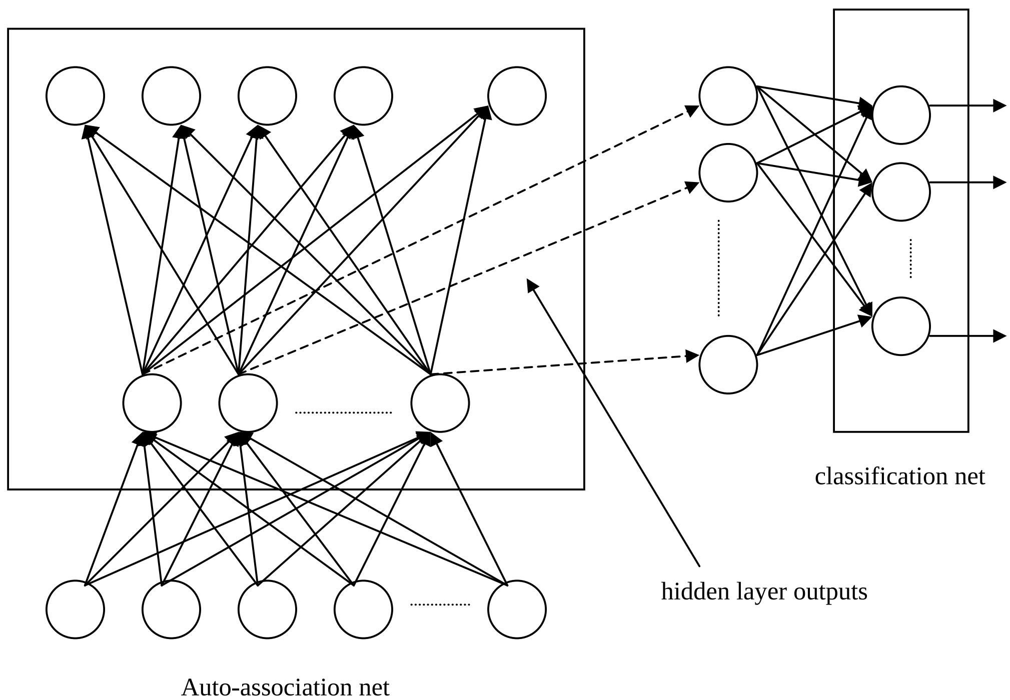


Figure 2.6: Auto-association and classification networks

Bourland and Kamp [108] showed that "under the best circumstances", when the sigmoidal functions at the network nodes are replaced by linear functions (when the network is linear), the feature vector is the same as that produced by the Karhunen-Loeve basis, or the eigenfaces. When the network is nonlinear, the feature vector could deviate from the best. The problem here turns out to be an application of the singular value decomposition.

Specifically, suppose that for each training face vector x_k (n-dimensional), $k=1, 2,...,N$, the outputs of the hidden layer and output layer for the auto-association net are h_k (p-dimensional, usually $p<<n$ and $p<N$) and y_k (n-dimensional), respectively, with

$$h_k = F(W_1 x_k), \quad y_k = W_2 h_k. \tag{2.37}$$

Here, W_1 (p by n) and W_2 (n by p) are corresponding weight matrixes and $F(.)$ is either linear or a nonlinear function, applied 'component by component'. If we pack x_k, y_k and h_k into matrixes as in the eigenface case, then above relations can be rewritten as

$$H = F(W_1 X), \quad Y = W_2 H. \tag{2.38}$$

Minimizing the training error for the auto-association net amounts to minimizing the Frobenius matrix norm

$$\|X - Y\|^2 = \sum_{k=1}^{n} \|x_k - y_k\|^2. \tag{2.39}$$

since $Y=W_2H$, its rank is no more than p. Hence, in order to minimize training error, $Y=W_2H$ should be the best rank-p approximation to X, which means

$$W_2 H = U_p \Lambda V_p^T \tag{2.40}$$

where

$U_p = [u_1, u_2, \ldots, u_p]^T$,

$V_p = [v_1, v_2, \ldots, v_p]^T$,

Are the first p left and right singular vectors in the SVD of X respectively, which also are the first p eigenvectors of XX^T and X^TX.

One way to achieve this optimum is to have a linear *F(.)* and to set the weights to

$$W_1^T = W_2 = U_p \tag{2.41}$$

Since U_p contains the first eigenvectors of XX^T, we have for any input x

$$h = W_1 x = U_p x \tag{2.42}$$

which is the same as the feature vector in the eigenface approach. However, it must be noted that the auto-association net, when it is trained by the BP algorithm with an nonlinear *F(.)*, generally can not achieve this optimal performance.

In [109], the first 50 principal components of the images are extracted and reduced to 5 dimensions using an auto-associative neural network. The resulting representation is classified using a standard multi-layer perceptron.

In a different approach, a hierarchical neural network, which is grown automatically and not trained with gradient descent, was used for face recognition by Weng and Huang [110].

The most successive face recognition with neural networks is a recent work of Lawrence et. al. [19] which combines local image sampling, a self organizing map neural network, and a convolutional neural network. In the corresponding work two different methods of representing local image samples have been evaluated. In each method, a window is scanned over the image. The first method simply creates a vector from a local window on the image using the intensity values at each point in the window. If the local window is a square of sides $2W+1$ long, centered on x_{ij}, then the vector associated with this window is simply $[x_{i-w,j-w}, x_{i-w,j-w+1},..., x_{ij},..., x_{i+w,j+w-1}, x_{i+w,j+w}]$. The second method creates a representation of the local sample by forming a vector out of the intensity of the center pixel and the difference in intensity between the center pixel and all other pixels within the square window. Then the vector is given by $[x_{ij}-x_{i-w,j-w}, x_{ij}-x_{i-w,j-w+1},..., w_{ij}x_{ij},..., x_{ij}-x_{i+w,j+w-1}, x_{ij}-x_{i+w,j+w}]$, $w_{i,j}$ is the weight of center pixel value $x_{i,j}$. The resulting representation becomes partially invariant to variations in intensity of the complete sample and the degree of invariance can be modified by adjusting the weightw_{ij}.

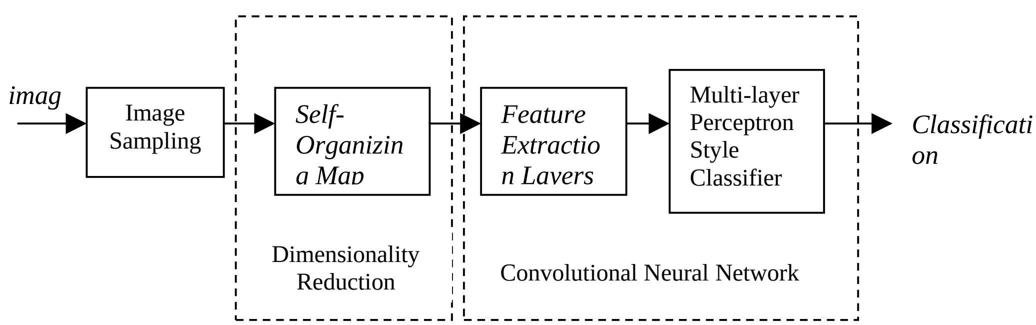


Figure 2.7: The diagram of the Convolutional Neural Network System

The self organizing map, SOM, introduced by Teudo Kohonen [111, 112], is an unsupervised learning process which learns the distribution of a set of patterns without any class information. The SOM defines a mapping from an input space R^n onto a topologically ordered set of nodes, usually in a lower dimensional space. For classification a convolutional network is used, as it achieves some degree of shift and deformation invariance using three ideas: local receptive field, shared weights, and spatial subsampling. The diagram of the system is shown in Figure 2.7.

2.2.6. Template Based Methods

The most direct of the procedures used for face recognition is the matching between the test images and a set of training images based on measuring the correlation. The matching technique is based on the computation of the normalized cross correlation coefficient C_N defined by,

$$C_N = \frac{E\{I_G I_T\} - E\{I_G\}E\{I_T\}}{\sigma\{I_T\}\sigma\{I_G\}} \tag{2.43}$$

Where I_G is the gallery image which must be matched to the test image, I_T. $I_G I_T$ is the pixel by pixel product, E is the expectation operator and σ is the standard deviation over the area being matched. This normalization rescales the test and gallery images energy distribution so that their variances and averages match. However correlation based methods are highly sensitive to illumination, rotation and scale changes. The best results for the reduction of the illumination changes were obtained using the intensity of gradient $\left(\left|\delta_X I_G\right| + \left|\delta_Y I_G\right|\right)$. Correlation method is computationally expensive, so the dependency of the recognition on the resolution of the image has been investigated.

In [16], Brunelli and Poggio describe a correlation based method for face recognition in which templates corresponding to facial features of relevant significance as the eyes, nose and mouth are matched. In order to reduce complexity, in this method first positions of those features are detected. Detection of facial features is also subjected to a lot of studies [36, 81, 113, 114]. The method purposed by Brunelli and Poggio uses a set of templates to detect the eye position in a new image, by looking for the maximum absolute values of the normalized correlation coefficient of these templates at each point in the test image. In order to handle scale variations, five eye templates at different scales were used. However, this method is computationally expensive, and also it must be noted that eyes of different people can be markedly different. Such difficulties can be reduced by using a hierarchical correlation [115].

After facial features are detected for a test face, they are compared to those of gallery faces returning a vector of matching scores (one per feature) computed through normalized cross correlation.

The similarity scores of different features can be integrated to obtain a global score. This cumulative score can be computed in several ways: choose the score of the most similar feature or sum the feature scores or sum the feature scores using weights. After cumulative scores are computed, a test face is assigned to the face class for which this score is maximized.

The recognition rate reported in [16] is higher than 96%. The correlation method as described above requires a robust feature detection algorithm with respect to variations in scale, illumination and rotations. Moreover the computational complexity of this method is quite high.

Beymer [116] extended the correlation based approach to a view based approach for recognizing faces under varying orientations, including rotations in depth. In the first stage three facial features (eyes and nose lobe) are detected to determine face pose. Although feature detection is similar to previously described correlation method to handle rotations, templates from different views and different people are used. After face pose is determined, matching procedure takes place with the corresponding view of the gallery faces. In this case, as the number of model views for each person in the database is increased, computational complexity is also increased.

2.2.7. Feature Based Methods

Since most face recognition algorithms are minimum distance classifiers in some sense, it is important to consider more carefully how a “distance” should be defined. In the previous examples (eigenface, neural nets… etc.) the distance between an observed face *x* and a gallery face *c* is the common Euclidean distance $d(x,c) = \|x - c\|$, and this distance is sometimes computed using an alternative orthonormal basis as $d(x,c) = \|\tilde{x} - \tilde{c}\|$.

While such an approach is easy to compute, it also has some shortcomings. When there is an affine transformation between two faces (shift and dilation), $d(x,c)$ will not be zero; in fact it can be quite large. As another example, when there is local transformations and deformations (x is a "smiling" version of c), again $d(x,c)$ will not be zero. Moreover, it is very useful to store information only about the key points of the face. Feature based approaches can be a solution to the above problems.

Manjunath et al. [36] proposed a method that recognizes faces by using topological graphs that are constructed from feature points obtained from Gabor wavelet decomposition of the face. This method reduces the storage requirements by storing facial feature points detected using the Gabor wavelet decomposition. Comparison of two faces begins with the alignment of those two graphs by matching centroids of the features. Hence, this method has some degree of robustness to rotation in depth but only under restrictedly controlled conditions. Moreover, illumination changes and occluded faces are not taken into account.

In the proposed method by Manjunaht et. al. [36], the identification process utilizes the information present in a topological graph representation of the feature points. The feature points are represented by nodes V_i $i=\{1, 2, 3,\ldots\}$, in a consistent numbering technique. The information about a feature point is contained in $\{S, q\}$, where S represents the spatial locations and q is the feature vector defined by,

$$q_i = \left[Q_i(x, y, \theta_1), \ldots, Q_i(x, y, \theta_N)\right] \quad (2.44)$$

corresponding to i^{th} feature point. The vector q_i is a set of spatial and angular distances from feature point i to its N nearest neighbors denoted by $Q_i(x,y,\theta_j)$, where j is the j^{th} of the N neighbors. N_i represents a set of neighbors. The neighbors satisfying both maximum number N and minimum Euclidean distance d_{ij} between two points V_i and V_j are said to be of consequence for i^{th} feature point.

In order to identify an input graph with a stored one, which might be different either in total number of feature points or in the location of the respective faces, two cost values are evaluated [36]. One is the topological cost and the other is a similarity cost. If i, j refer to nodes in the input graph I and x', y', m', n' refer to nodes in the stored graph O then the two graphs are matched as follows [36]:

1. The centroids of the feature points of I and O are aligned.
2. Let N_i be the i^{th} feature point $\{V_i\}$ of I. Search for the best feature point $\{V_{i'}\}$ in O using the criterion

$$S_{ii'} = 1 - \frac{q_i . q_{i'}}{\|q_i\| \|q_{i'}\|} = \min_{m' \in N_i} S_{im'} \tag{2.45}$$

3. After matching, the total cost is computed taking into account the topology of the graphs. Let nodes i and j of the input graph match nodes i' and j' of the stored graph and let $j \in N_i$ (i.e., V_j is a neighbor of V_i). Let $\rho_{ii'jj'} = \min\left\{\frac{d_{ij}}{d_{i'j'}}, \frac{d_{i'j'}}{d_{ij}}\right\}$. The topology cost is given by

$$\mathrm{T}_{ii'jj'} = 1 - \rho_{i'jj'}. \tag{2.46}$$

4. The total cost is computed as

$$C_1 = \sum_i S_{ii'} + \lambda_t \sum_i \sum_{j \in N_i} \mathrm{T}_{ii'jj'} \tag{2.47}$$

where λ_t is a scaling parameter assigning relative importance to the two cost functions.

5. The total cost is scaled appropriately to reflect the possible difference in the total number of the feature points between the input and the stored graph. If n_i, n_o are the numbers of the feature points in the input and stored graph, respectively, then scaling factor $s_f = \max\left\{\frac{n_I}{n_O}, \frac{n_O}{n_I}\right\}$ and the scaled cost is $C(I,O)=s_f C_1(I,O)$.

6. The best candidate is the one with the least cost,

$$C(I,O^*) = \min_{O'} C(I,O') \tag{2.48}$$

The recognized face is the one that has the minimum of the combined cost value. In this method [36], since comparison of two face graphs begins with centroid alignment, occluded cases will cause a great performance decrease. Moreover, directly using the number of feature points of faces can be result in wrong classifications while the number of feature points can be changed due to exterior factors (glasses…etc).

Another feature based approach is the elastic matching algorithm proposed by Lades et al. [117], which has roots in aspect-graph matching. Let S be the original two-dimensional image lattice (Figure 2.8). The face template is a vector field by defining a new type of representation,

$$c=\{c_i, i \in S_1\} \tag{2.49}$$

where S_1 is a lattice embedded in S and c_i is a feature vector at position i. S_1 is much coarser and smaller than S. c should contain only the most critical information about the face, since c_i is composed of the magnitude of the Gabor filter responses at position $i \in S_1$. The Gabor features c_i provide multi-scale edge strengths at position i.

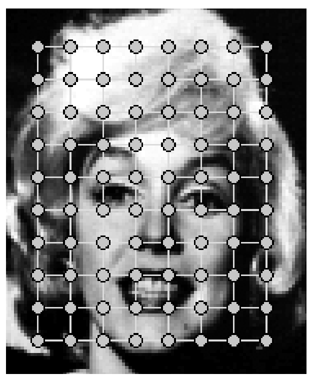


Figure 2.8: A 2D image lattice (grid graph) on Marilyn Monroe's face.

An observed face image is defined as a vector field on the original image lattice S.

$$x=\{x_j, j \in S\} \tag{2.50}$$

where x_j is the same type of feature vector as c_i but defined on the fine-grid lattice S.

In the elastic graph matching approach [23, 27, 117], the distance $d(c,x)$ is defined through 'best match' between x and c. A match between x and c can be described uniquely through a mapping between S_1 and S.

$$M{:}S_1 \rightarrow S \tag{2.51}$$

However without restriction, the total number of such mappings is $\|S\|^{\|S_1\|}$.

Recognizing a face, the best match should preserve both features and local geometry. If $i \in S_1$ and $j=M(i)$, then feature preserving means that x_j is not much different from c_i. In addition, if i_1 and i_2 are close in S_1 then preserving local geometry means $j_1=M(i_1)$ to be close to $j_2=M(i_2)$. Such a match is called *elastic* since the preservation can be approximate rather than exact, the lattice S_1 can be stretched unevenly.

Finding the best match is based on minimizing the following energy function [117],

$$E(M) = \sum_i \left[-\frac{\langle c_i, x_j \rangle}{\|c_i\| \|x_j\|} \right] + \lambda \sum_{i_1, i_2} \left[(i_1 - i_2) - (j_1 - j_2) \right]^2. \tag{2.52}$$

Due to the large number of possible matches, this might be difficult to obtain in an acceptable time. Hence, an approximate solution has to be found. This is achieved in two stages: Rigid matching and deformable matching [117]. In rigid matching c is moved around in x like conventional template matching and at each position $\|c - x'\|$ is calculated. Here x' is the part of x that, after being matched with c, does not lead to any deformation of S_1. In deformable matching, lattice S_1 is stretched through random local perturbations to further reduce the energy function.

There is still a question of face template construction. An automated scheme has been shown to work quite well. The procedure is as follows [117]:

1. Pick the face image of an individual and generate a face vector field using Gabor filters. Denote this as vector field *x*.
2. Place a coarse grid lattice S_1 ‘by hand’ on x such that vertices are close to important facial features such as the eyes.
3. Collect the feature vectors at vertices of S_1 to form the template c.
4. Pick another face image (of a different person) and generate a face vector field, denoted as *y*.
5. Perform template matching between the *c* with *y* using rigid matching.
6. When a best match is found, collect the feature vectors at vertices of S_1 to form the template of *y*.
7. Repeat steps 4-6.

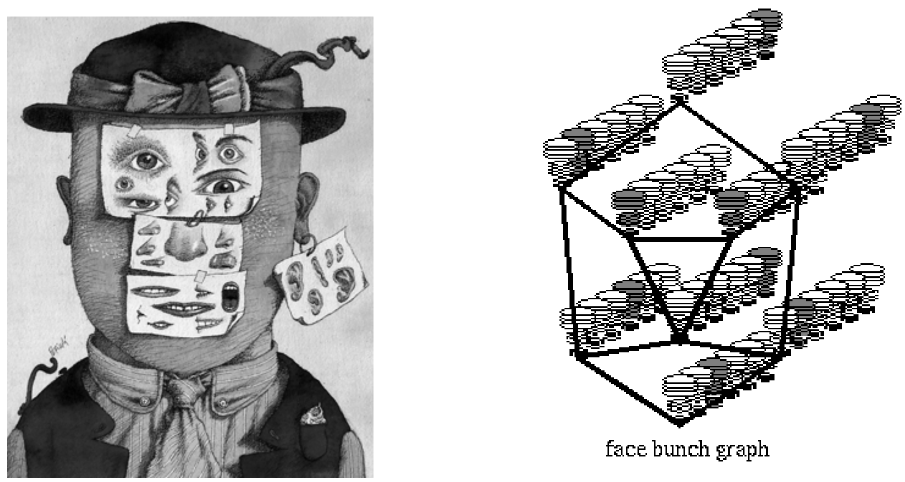
Malsburg et al. [23, 27] developed a system based on graph matching approach on with several major modifications. First, they suggested using not only magnitude but also phase information of the Gabor filter responses. Due to phase rotation, vectors taken from image points only a few pixels apart from each other have very different coefficients, although representing almost the same local feature. Therefore, phase should either be ignored or compensated for its variation explicitly.

As mentioned in [23], using phase information has two potential advantages,

1. Phase information is required to discriminate between patterns with similar magnitudes.
2. Since phase varies quickly with location provides a means for accurate vector localization in an image.

Malsburg et al. proposed to compensate phase shifts by estimating small relative displacements between two feature vectors to use a phase sensitive similarity function [23].

Second modification is the use of bunch graphs (Figure 2.9, 2.10). The face bunch graph is able to represent a wide variety of faces, which allows matching on face images of previously unseen individuals. These improvements make it possible to extract an image graph from a new face image in one matching process. Computational efficiency, and ability to deal with different poses explicitly are the major advantages of the system in [23], compared to [117].



(a) (b)

Figure 2.9: A brunch graph from the a) artistic point of view (*``Unfinished Portrait''* by *Tullio Pericoli* (1985)), b) scientific point of view

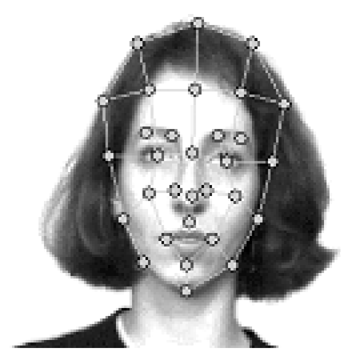


Figure 2.10: Bunch graph matched to a face.

2.2.8. Current State of the Art

Comparing recognition results of different face recognition systems is a complex task, because generally experiments are carried out on different datasets. However, the following 5 questions can help while reviewing different face recognition methods:

1. Were expression, head orientation and lighting conditions controlled?
2. Where subjects allowed wearing glasses and having beards or other facial masks?
3. Was the subject sample balanced? Where gender, age and ethnic origin spanned evenly?
4. How many subjects there in database? How many images were used for training and testing?
5. Were the face features located manually?

Answering the above questions contributes to building better description of the constraints within each approach operated. This helps to make a more fair comparison between different set of experimental results. However the most direct and reliable comparison between different approaches is obtained by experimenting with the same database.

By 1993, there were several algorithms claiming to have accurate performance in minimally consternated environments. For a better comparison of those algorithms DARPA and Army Research Laboratory established the FERET program (see section 4.1.4) with the goals of both evaluating their performance and encouraging advances in the technology [83].

Today, there are three algorithms that have demonstrated the highest level of recognition accuracy on large databases (1196 people or more) under double blind testing conditions. These are the algorithms from University of Southern California (USC), University of Maryland (UMD), and MIT Media Lab [83, 128]. The MIT, Rockefeller and UMD algorithms all use a version of the eigenface transform followed by discriminative modeling (section 2.2.3.1.2). However the USC system uses a very different approach. It begins by computing Gabor Wavelet transform of the image and does the comparison between images using a graph matching algorithm (section 2.2.7). Only two of these algorithms, from USC and MIT are capable of both minimally constrained detection and recognition; the others require approximate eye locations to operate. Algorithm developed at Rockefeller University, was an early contender, dropped from testing to form a commercial enterprise. The MIT and USC algorithms have also become the basis for commercial systems. In FERET testing, the performance of the four algorithms is similar enough that it is difficult or impossible to make meaningful distinctions between them. Results of FERET test will be given at section 4.1.4.

CHAPTER 3

FACE REPRESENTATION AND MATCHING USING GABOR WAVELETS

Using local features is a mature approach to face recognition problem [11, 14, 17, 18, 23, 35, 59, 36, 62]. One of the main motivations of feature based methods is due to: representation of the face image in a very compact way and hence lowering the memory needs. This fact especially gains importance when there is a huge face database. Feature based methods are based on finding fiducial points (or local areas) on a face and representing corresponding information in an efficient way. However, choosing suitable feature locations and the corresponding values are extremely critical for the performance of a recognition system. Searching nature for finding an answer has lead researchers to examine the behavior of human visual system (HVS).

Physiological studies found simple cells, in human visual cortex, that are selectively tuned to orientation as well as to spatial frequency. It was suggested that the response of a simple cell could be approximated by 2 D Gabor filters

[119]. Over the last couple of years, it has been shown that using Gabor filters as the front-end of an automated face recognition system could be highly successful [23, 36, 35, 27, 32]. One of the most successful face recognition method is based on graph matching of coefficients which are obtained from Gabor filter responses [83, 23]. However, such graph matching algorithm methods have some disadvantages due to their matching complexity, manual localization of training graphs, and overall execution time. They use general face structure to generate graphs and such an approach brings the question of how efficient the feature represents the special facial characteristics of each individual. A novel Gabor based method may overcome those disadvantages.

2 D Gabor functions are similar to enhancing edge contours, as well as valleys and ridge contours of the image. This corresponds to enhancing eye, mouth, nose edges, which are supposed to be the main important points on a face. Moreover, such an approach also enhances moles, dimples, scars, etc. Hence, by using such enhanced points as feature locations, a feature map for each facial image can be obtained and each face can be represented with its own characteristics without any initial constrains. Having feature maps specialized for each face makes it possible to keep overall face information while enhancing local characteristics.

In this thesis, a novel method is proposed based on selecting peaks (high-energized points) of the Gabor wavelet responses as feature points, instead of using predefined graph nodes as in elastic graph matching [23] which reduces the representative capability of Gabor wavelets. Feature vectors are constructed by sampling Gabor wavelet transform coefficients at feature points.

In the following sections, the details about the Gabor wavelet transform will be presented while giving the reasons of using it for face recognition. Then the proposed algorithm will be explained, explicitly.

3.1. Face Representation Using Gabor Wavelets

3.1.1. Gabor Wavelets

Since the discovery of crystalline organization of the primary visual cortex in mammalian brains thirty years ago by Hubel and Wiesel [121], an enormous amount of experimental and theoretical research has greatly advanced our understanding of this area and the response properties of its cells. On the theoretical side, an important insight has been advanced by Marcelja [122] and Daugman [118, 119] that simple cells in the visual cortex can be modeled by Gabor functions. The Gabor functions proposed by Daugman are local spatial bandpass filters that achieve the theoretical limit for conjoint resolution of information in the 2D spatial and 2D Fourier domains.

Gabor functions first proposed by Dennis Gabor as a tool for signal detection in noise. Gabor [120] showed that there exists a "quantum principle" for information; the conjoint time-frequency domain for 1D signals must necessarily be quantized so that no signal or filter can occupy less than certain minimal area in it. However, there is a trade off between time resolution and frequency resolution. Gabor discovered that Gaussian modulated complex exponentials provide the best trade off. For such a case, the original Gabor elementary functions are generated with a fixed Gaussian, while the frequency of the modulating wave varies.

Gabor filters, rediscovered and generalized to 2D, are now being used extensively in various computer vision applications. Daugman [118, 119] generalized the Gabor function to the following 2D form in order to model the receptive fields of the orientation selective simple cells:

$$\Psi_i(\vec{x}) = \frac{\left\|\vec{k}_i\right\|^2}{\sigma^2} e^{-\frac{\left\|\vec{k}_i\right\|^2 \|\vec{x}\|^2}{2\sigma^2}} \left[e^{j\vec{k}_i \vec{x}} - e^{-\frac{\sigma^2}{2}} \right] \tag{3.1}$$

Each ψ_i is a plane wave characterized by the vector $\vec{k}_i$ enveloped by a Gaussian function, where σ is the standard deviation of this Gaussian. The center frequency of i^{th} filter is given by the characteristic wave vector,

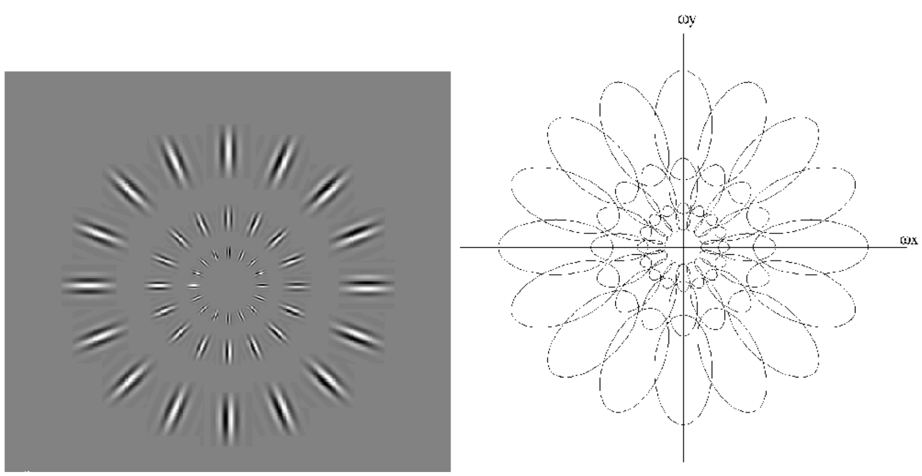
$$\vec{k}_i = \begin{pmatrix} k_{ix} \\ k_{iy} \end{pmatrix} = \begin{pmatrix} k_v \cos\theta_\mu \\ k_v \sin\theta_\mu \end{pmatrix}, \tag{3.2}$$

having a scale and orientation given by (k_v, θ_μ). The first term in the brackets (3.1) determines the oscillatory part of the kernel, and the second term compensates for the DC value of the kernel. Subtracting the DC response, Gabor filters becomes insensitive to the overall level of illumination.

Recent neurophysiological evidence suggests that the spatial structure of the receptive fields of simple cells having different sizes is virtually invariant. Daugman [118] and others [121, 122] have proposed that an ensemble of simple cells is best modeled as a family of 2D Gabor wavelets sampling the frequency domain in a log-polar manner. This class is equivalent to a family of affine coherent states generated by rotation and dilation. The decomposition of an image *I* into these states is called the *wavelet* transform of the image:

$$R_i(\vec{x}) = \int I(\vec{x}')\Psi_i(\vec{x} - \vec{x}')d\vec{x}' \tag{3.3}$$

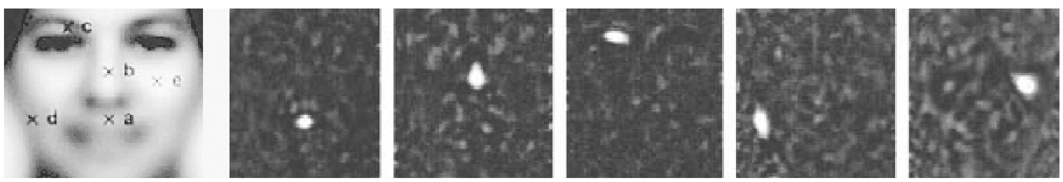
Where $I(\vec{x})$ is the image intensity value at $\vec{x}$.



(a) (b)

Figure 3.1: (a) an ensemble of Gabor wavelets, (b) their coverage of spatial frequency plane

Each member of this family of Gabor wavelets models the spatial receptive field structure of a simple cell in the primary visual cortex. The Gabor decomposition can be considered as a *directional microscope* with an orientation and scaling sensitivity. Due to the end-inhibition property of these cells, they response to short lines, line endings and sharp changes in curvature. Since such curves correspond to some low-level salient features in an image, these cells can be assumed to form a low level feature map of the intensity image (Figure 3.2).



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

Figure 3.2: Small set of features can recognize faces uniquely, and receptive fields that are matched to the local features of the face (a) mouth, (b) nose, (c) eyebrow, (d)jawline, (e) cheekbone.

Since the Gabor wavelet transform is introduced to computer vision area, one of the most important application areas for 2D Gabor wavelet representation is face recognition (see Section 2.2.4). In a U.S. government activity (FERET program) to find the best face recognition system, a system based on Gabor wavelet representation of the face image performed among other systems on several tests. Although the recognition performance of this system shows qualitative similarities to that of humans by now means, it still leaves plenty of room for improvement.

Utilization of the 2D Gabor wavelet representation in computer vision was pioneered by Daugman in the 1980s. More recently, B. S. Manjunath [36] et al has developed a face recognition system based on this representation. Afterwards, studies for Gabor wavelet representation in the field of face recognition by using is continued with appending dynamic link architecture [117] and elastic graph matching [23] to the previous system.

3.1.2. 2D Gabor Wavelet Representation of Faces

Since face recognition is not a difficult task for human beings, selection of biologically motivated Gabor filters is well suited to this problem. Gabor filters, modeling the responses of simple cells in the primary visual cortex, are simply plane waves restricted by a Gaussian envelope function (3.1) [12].

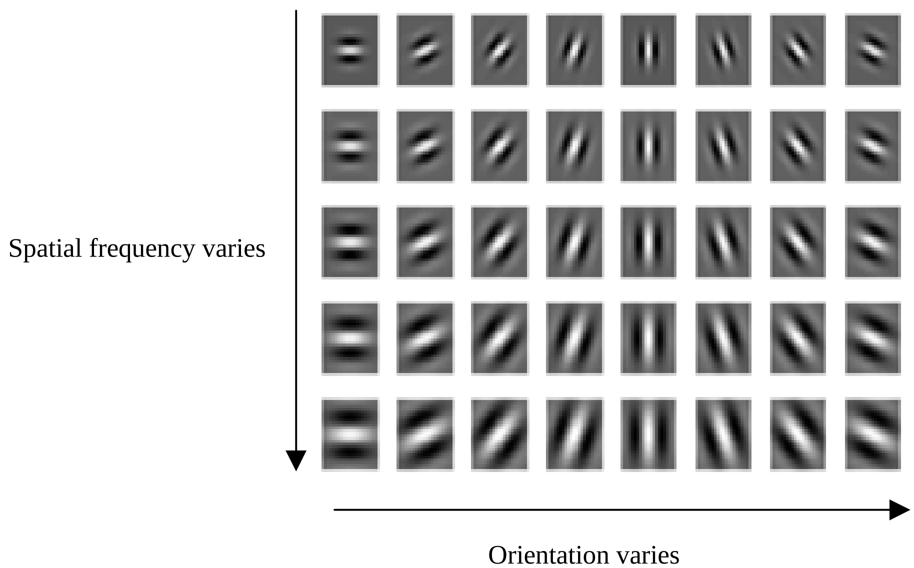
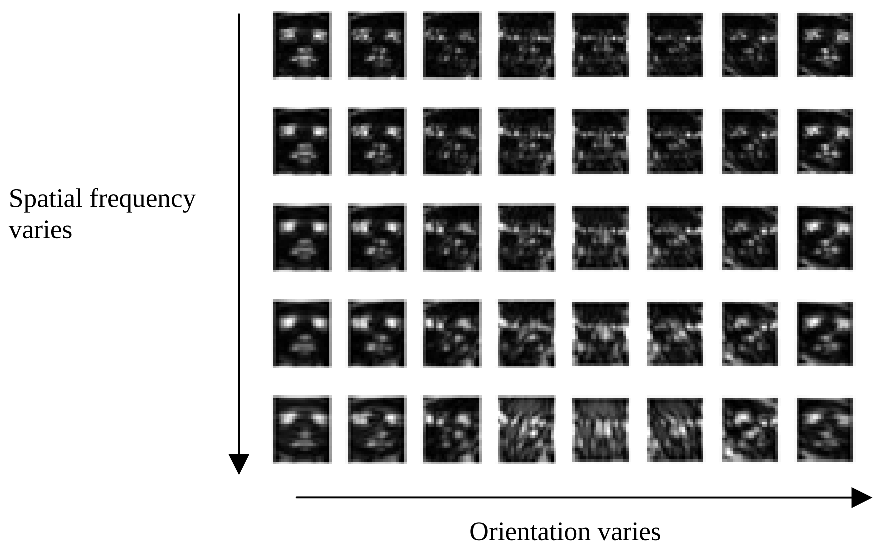


Figure 3.3: Gabor filters correspond to 5 spatial frequency and 8 orientation.

An image can be represented by the Gabor wavelet transform allowing the description of both the spatial frequency structure and spatial relations. Convolving the image with complex Gabor filters with 5 spatial frequency (v = 0,…,4) and 8 orientation (μ = 0,…,7) captures the whole frequency spectrum, both amplitude and phase (Figure 3.3). In Figure 3.4, an input face image and the amplitude of the Gabor filter responses are shown.

(a)



(b)

Figure 3.4: Example of a facial image response to above Gabor filters, a) original face image (from Stirling database), and b) filter responses.

One of the techniques used in the literature for Gabor based face recognition is based on using the response of a grid representing the facial topography for coding the face. [23, 25, 26, 35]. Instead of using the graph nodes, high-energized points can be used in comparisons which forms the basis of this work. This approach not only reduces computational complexity, but also improves the performance in the presence of occlusions.

3.1.3. Feature extraction

Feature extraction algorithm for the proposed method has two main steps (Figure 3.6): (1) Feature point localization, (2) Feature vector computation.

3.1.3.1. Feature point localization

In this step, feature vectors are extracted from points with high information content on the face image. In most feature-based methods, facial features are assumed to be the eyes, nose and mouth. However, we do not fix the locations and also the number of feature points in this work. The number of feature vectors and their locations can vary in order to better represent diverse facial characteristics of different faces, such as dimples, moles, etc., which are also the features that people might use for recognizing faces (Figure 3.5).

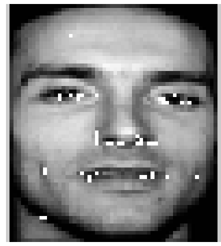


Figure 3.5: Facial feature points found as the high-energized points of Gabor wavelet responses.

From the responses of the face image to Gabor filters, peaks are found by searching the locations in a window W_0 of size *WxW* by the following procedure: A feature point is located at (x_0, y_0), if

$$R_j(x_0, y_0) = \max_{(x,y)\in W_0}(R_j(x, y)) \tag{3.4}$$

$$R_j(x_0, y_0) > \frac{1}{N_1 N_2}\sum_{x=1}^{N_1}\sum_{y=1}^{N_2} R_j(x, y), \tag{3.5}$$

$j=1,\ldots,40$

where R_j is the response of the face image to the j^{th} Gabor filter (3.3). $N_1\, N_2$ is the size of face image, the center of the window, W_0 is at (x_0, y_0). Window size W is one of the important parameters of proposed algorithm, and it must be chosen small enough to capture the important features and large enough to avoid redundancy Equation (3.5) is applied in order not to get stuck on a local maximum, instead of finding the peaks of the responses. In our experiments a *9x9* window is used to search feature points on Gabor filter responses. A feature map is constructed for the face by applying above process to each of 40 Gabor filters.

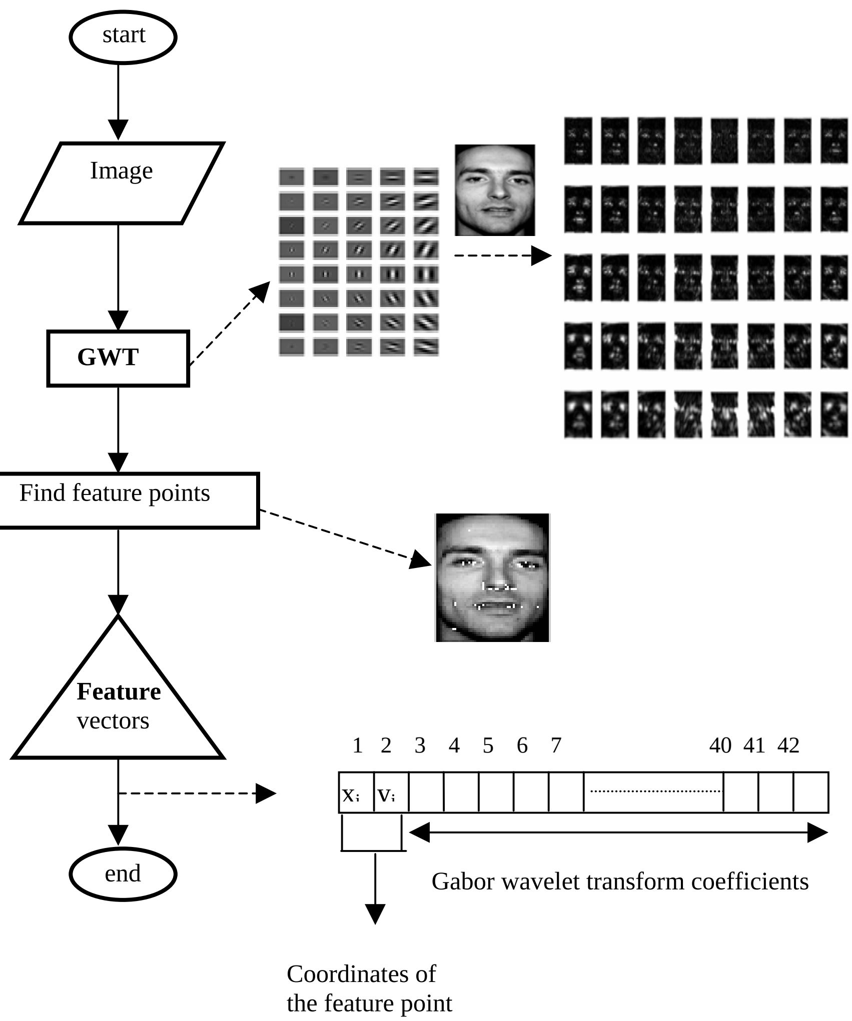


Figure 3.6: Flowchart of the feature extraction stage of the facial images.

3.1.3.2. Feature vector generation

Feature vectors are generated at the feature points as a composition of Gabor wavelet transform coefficients. k^{th} feature vector of i^{th} reference face is defined as,

$$v_{i,k} = \{x_k, y_k, R_{i,j}(x_k, y_k) \; j = 1,.....,40\}. \tag{3.7}$$

While there are 40 Gabor filters, feature vectors have 42 components. The first two components represent the location of that feature point by storing *(x, y)* coordinates. Since we have no other information about the locations of the feature vectors, the first two components of feature vectors are very important during matching (comparison) process. The remaining 40 components are the samples of the Gabor filter responses at that point.

Although one may use some edge information for feature point selection, here it is important to construct feature vectors as the coefficients of Gabor wavelet transform. Feature vectors, as the samples of Gabor wavelet transform at feature points, allow representing both the spatial frequency structure and spatial relations of the local image region around the corresponding feature point.

3.2. Matching Procedure

3.2.1. Similarity Calculation

In order to measure the similarity of two complex valued feature vectors, following similarity function is used which ignores the phase:

$$S_i(k, j) = \frac{\sum_l |v_{i,k}(l)|\,|v_{t,j}(l)|}{\sqrt{\sum_l |v_{i,k}(l)|^2 \sum_l |v_{t,j}(l)|^2}}, \quad l = 3,\ldots,42. \qquad (3.8)$$

$S_i(k,j)$ represents the similarity of j^{th} feature vector of the test face, $(v_{i,j})$, to k^{th} feature vector of i^{th} reference face, $(v_{i,k})$, *where l* is the number of vector elements. Proposed similarity measure between two vectors satisfies following constrains:

$0 < S_i < 1$,

and if i^{th} gallery face image is used also as the test image,

$S_i(j, j) = 1$.

The location information is not used for vector similarity calculation, but only the magnitudes of the wavelet coefficients are take place at (3.8). It must be clarified that the similarity function (3.8) is only one component of the proposed matching procedure (Section 3.2.2). Location information of feature vectors will also be used during matching.

Equation (3.8) is a very common similarity measure between feature vectors, containing Gabor wavelet transform coefficients [36], but sometimes we might have small variations [23, 27]. In [23] similarity function at (3.8) is used with complex valued coefficients and an additional phase compensating term. In the early experiments it is observed that small spatial displacements cause change in complex valued coefficients due to phase rotation. Then phase can either be ignored or compensated as in [23]. Although phase compensated similarity function is found to increase recognition performance significantly [23,27], similarity function without phase is chosen to avoid computational complexity.

3.2.2. Face comparison

After feature vectors are constructed from the test image, they are compared to the feature vectors of each reference image in the database. This comparison stage takes place in two steps. In the first step, we eliminate the feature vectors of the reference images which are not close enough to the feature vectors of the test image in terms of location and similarity. Only the feature vectors that fit the following two criterions are examined in the next step.

1. $\sqrt{(x_r - x_t)^2 + (y_r - y_t)^2} < th_1$,

   where th_1 is the approximate radius of the area that contains either eye, mouth or nose, (x_r, y_r) and (x_t, y_t) represents the location of a feature point on a reference face and test face respectively. Comparing the distances between the coordinates of the feature points simply avoids the matching of a feature point located around the eye with a point of a reference facial image that is located around the mouth. After such a localization, we may disregard the location information in the second step. Moreover here topology of face is also examined to use corresponding information at the final matching by only letting feature points that are match each other in a topological manner.

2. $S_i(k,j)>th_2,$

   Similarity of two feature vectors is greater than *th2*, where th_2 is chosen as the standard deviation of similarities of all feature vectors in the reference gallery and the similarity of two vectors is computed by Equation (3.8).

   Although this thresholding seems to be done very roughly it reduces the number of feature vectors at the gallery faces and increases the speed of algorithm at the following steps.

By changing th_1 and th_2 one can control the topology and vector similarity costs. In other words, increasing th_1 gives more area for searching the feature points with similarities larger than th_2. This can be useful when the locations of features changed due to some reasons, such as different expression. However, if th_1 is too large then the face topology information could be totally wrong. By keeping th_1 constant, increasing th_2 in a large extent will result in finding no match, and conversely decreasing th_2 could result in a redundant feature vector that will increase the computational cost. However, small variations in th_1 and th_2 will not effect the performance of the method. One can choose these values stated at above steps 1 and 2.

As a result of the first step, we are left with N_k feature vectors of the reference face and N_{kj} of them are known to be close enough in terms of both location similarity to the j^{th} feature vector of the test face. Hence, possible matches are determined to each of the feature vector on the test face from the feature vectors of gallery faces.

After elimination of feature vectors in the gallery, at the end of the first step, there could be no remaining feature vector from the gallery faces to be matched to a feature vector of the test face. When such a case occurs, that feature vector of test face is ignored and matching procedure is continued with others.

In the second step, Equation (3.9) is applied as a second elimination of feature vectors of the reference face in order to guarantee surviving at most only one feature vector of a reference face to be matched with a feature vector of the test face:

$$Sim_{i,j} = \max_{l \in N_{k,j}} (S_i(l, j)), \tag{3.9}$$

In Equation (3.9) $Sim_{i,j}$ gives the similarity of the i^{th} reference face to the test face based on the j^{th} feature vector.

Eventually, the overall similarity of each reference face is computed as the mean of feature vector similarities that passed the two steps (3.10).

$$OS_i = mean\{Sim_{i,j}\}. \tag{3.10}$$

OS_i represents the overall similarity of test face to i^{th} reference face, and its values change from 0 to 1. If i^{th} gallery image used also as the test image, then OS_i will be equal to 1.

Although *OS* gives a good measure for face similarity, it can further be improved by considering the number of feature vectors. It should be noted that the number of feature points for any of two face image is not equal even for images from the same individual. The reason is due to the huge variation of the number of feature vectors (e.g. glasses/no glasses, illumination changes, etc.) from face to face and also for the same face under different conditions. Since the number of feature vectors that give the overall similarity (3.10) is not determined by *OS,* it is not very meaningful to use it alone. As an example, *OS=0.85* as the mean of 20 vector similarities should be more valuable than *OS=0.95* as the mean of 2 vector similarities. Moreover, numbers of matched feature vectors of each gallery face gives an information about the topological matching. In order to emphasize the information contained by the number of matched feature points, a new parameter *C* is computed, by only counting the number of feature vectors of a gallery face that have high similarity to a feature vector of the test face:

$$C_i = \sum_j \delta(Sim_{i,j} = \max(Sim_{l,j})), \tag{3.11}$$

$$\sum_l C_l = N_t , \tag{3.12}$$

l=1,…, number of reference faces

where N_t is the total number of feature vectors of test face, and $\delta(.)$ is the Delta Dirac function.

For each feature vector of test face we sort reference faces by means of similarity, and count the number of times that each reference face gets the first place (3.11).

The simulation results are shown in Figure 3.7. In this representation for ideal results, the corresponding figure must be a stair case function whose horizontal width includes the test image for each person in the gallery and vertical level is the correct result. Recognition results achieved by both comparing only similarities (Figure 3.7-a) and comparing only number of times having maximum similar feature vector (Figure 3.7-b) are presented in Figure 3.7.

Although results achieved in Figure 3.7-b are better, a problem occurs when a test face has more feature vectors (due to illumination changes, having glasses…etc.) than the corresponding reference face. We can explain this better by an example: Assume that there are two reference faces *i* and *j*, and a test face, having the number of feature vectors N_i, N_j and N_t respectively. i^{th} reference face reaches the maximum similarity at its all feature points, $(C_i=N_i)$. Having more feature points j^{th} reference face gets maximum similarity at more points than i^{th} does, $(C_i < C_j < N_t)$ and $(C_j<N_j)$. In the case, one can use the number of feature vectors of reference faces as an additional criteria.

In order to find the best match, the weighted sum of these two results can be used. This improves the performance, in case different false matches are obtained by using individual results of *OS* and *C*. However, successive matching can be done by using the expectation that if i^{th} reference face is correct it should get high values for both C_i and OS_i.

Hence, the best candidate match is searched to maximize the following function,

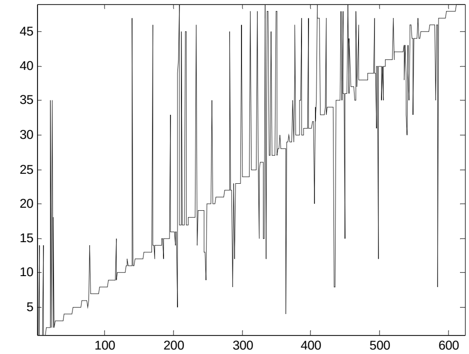
$$FSF_i = \alpha\, OS_i + \beta \left(\frac{C_i}{N_i} \right). \qquad (3.13)$$

where C_i is the number of feature vectors of i^{th} reference face that have maximum similarity to a feature vector of test face, N_i is the number of feature vectors of i^{th} reference image. If i^{th} gallery image used also as the test image FSF_i will be equal to unity.

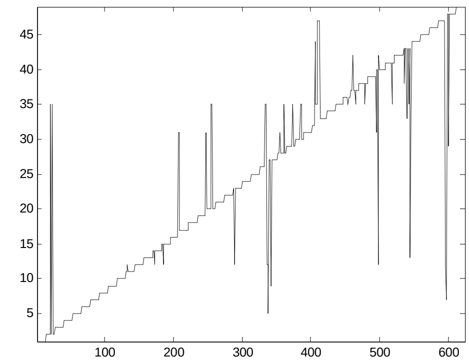
Although for the corresponding database %100 recognition result is achieved (Figure 3.7-c), it is seen that for larger databases counting only the maximum similar feature vectors of gallery faces as in (3.11) becomes useless. Instead, Equation (3.11) can be generalized for large databases by counting the number of feature vectors of each gallery face which is in the first %10 of similarity rank,

$$C_i = \sum_j \delta(Sim_{i,j} = rank_{10}(Sim_{l,j})), \tag{3.14}$$

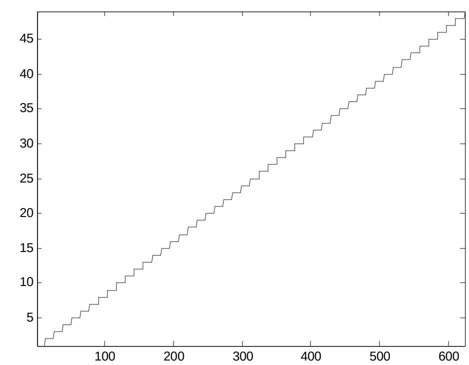
*l=1,…,*number of reference faces .



(a)



(b)



(c)

Figure 3.7: Test faces (1-624) vs. matching face from gallery (1-48) by comparing, a) only similarities, b) only number of maximum similar feature vectors, c) both similarities and number of maximum similar feature vectors.

CHAPTER 4

RESULTS

4.1. Simulation Setup

The proposed method is tested on four different face databases: Stirling [28], Purdue [29], ORL [30] and FERET [128] face databases. For each dataset, one frontal face image with neutral pose of each individual is placed in the gallery and the others are placed in the probe (test) set. Each database more than one face image with different conditions (expression, illumination…etc.), of each individual. It must be noted that none of the gallery images is found in the probe set. Each probe image is matched against the data in the gallery, and the ranked matches are analyzed.

Each of the above mentioned databases have different characteristics to test the proposed algorithm. Stirling face database [28] is used to measure the robustness against the expression changes. Purdue face database [29] contains occluded cases (some obstacles on face). Head pose variations and images that were taken at different times have been included in ORL face database [30]. Finally FERET face database [128] is used. FERET is not only a large face database (1196 individuals); but also challenging benchmark of all face recognition systems. This gives us an opportunity to compare the performance of our face recognition method by the others using a standardized dataset and test procedure.

In the following section, detailed information for those four face databases and their corresponding performance results for the proposed face recognition method are given with the comparisons with some major face recognition methods.

(a) (b)

Figure 4.1: Examples of different facial expressions of two people from Stirling database, a) gallery faces, b) probe faces.

4.1.1. Results for University of Stirling face database

University of Stirling face database contains grayscale facial images of 35 people (18 female, 17 male), in 3 poses and 3 expressions with constant illumination conditions (Figure 4.1). In the experiments, we used three frontal images with different expressions for each of 35 persons. The neutral views are placed in the gallery, whereas two other expressions (smiling and speaking) are used as probes. In this database the proposed algorithm gives 100% correct recognition. We did not reach any reported performance results on this database.

4.1.2. Results for Purdue University face database

Purdue University face database was created by Alex Martinez and Robert Benavente in the Computer Vision Center (CVC) at the U.A.B. It contains over 4,000 color images corresponding to 126 people's faces (70 men and 56 women). Images consist of frontal view faces with different facial expressions, illumination conditions, and occlusions (sunglasses and scarf). The pictures were taken at the CVC under strictly controlled conditions. No restrictions on wear (clothes, glasses, etc.); make-up, hairstyle, etc. were imposed to participants.

We have used images of first 58 (18 female, 40 male) individuals in our test setup due to some availability problems of database. For each individual, facial image with neutral expression is placed in the gallery and the remaining 12 with different conditions are used as probe. An example of 13 different facial images for a typical person from Purdue database is shown in Figure 4.2. After simulations, it is observed that 100% of people are correctly classified. Even none of the occluded cases are included, recognition performances of eigenface and a related eigenhills method for this database is reported as 82.3, 89.4 respectively [31]. Eigenhills and eigenfaces methods were highly effected by illumination changes, however proposed method is more robust to illumination changes as a property of Gabor wavelets (Section 3.1.1). Moreover as a result of using local distinct features, instead of a face template or local features bounded by a graph, proposed method gives a high performance result on occluded cases. In the proposed algorithm the facial features are compared locally, instead of using a general structure, hence it allows us to make a decision from the parts of the face. For example, when there are sunglasses, the algorithm compares faces in terms of mouth, nose and any other features rather than eyes.

Since the simulation results on this database are encouraging, we pass to ORL database on which simulation results of other wellkonwn face recognition methods are reported.

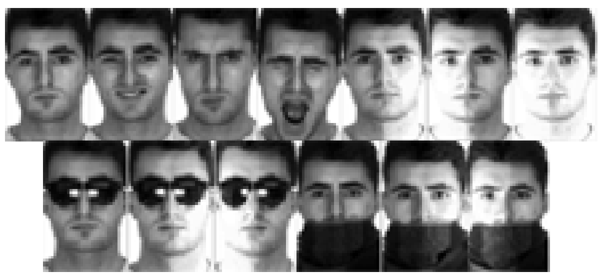


Figure 4.2: Example of different facial images for a person from Purdue database; first image is used for gallery and the rest 12 are for probe set.

Method	**Recognition rate (%)**
Eigenface [20]	82.3
Eigenhills [31]	89.4
Proposed face recognition method using GWT	100.0

Table 4.1: Recognition performances of eigenface, eigenhills and proposed method on the Purdue face database.

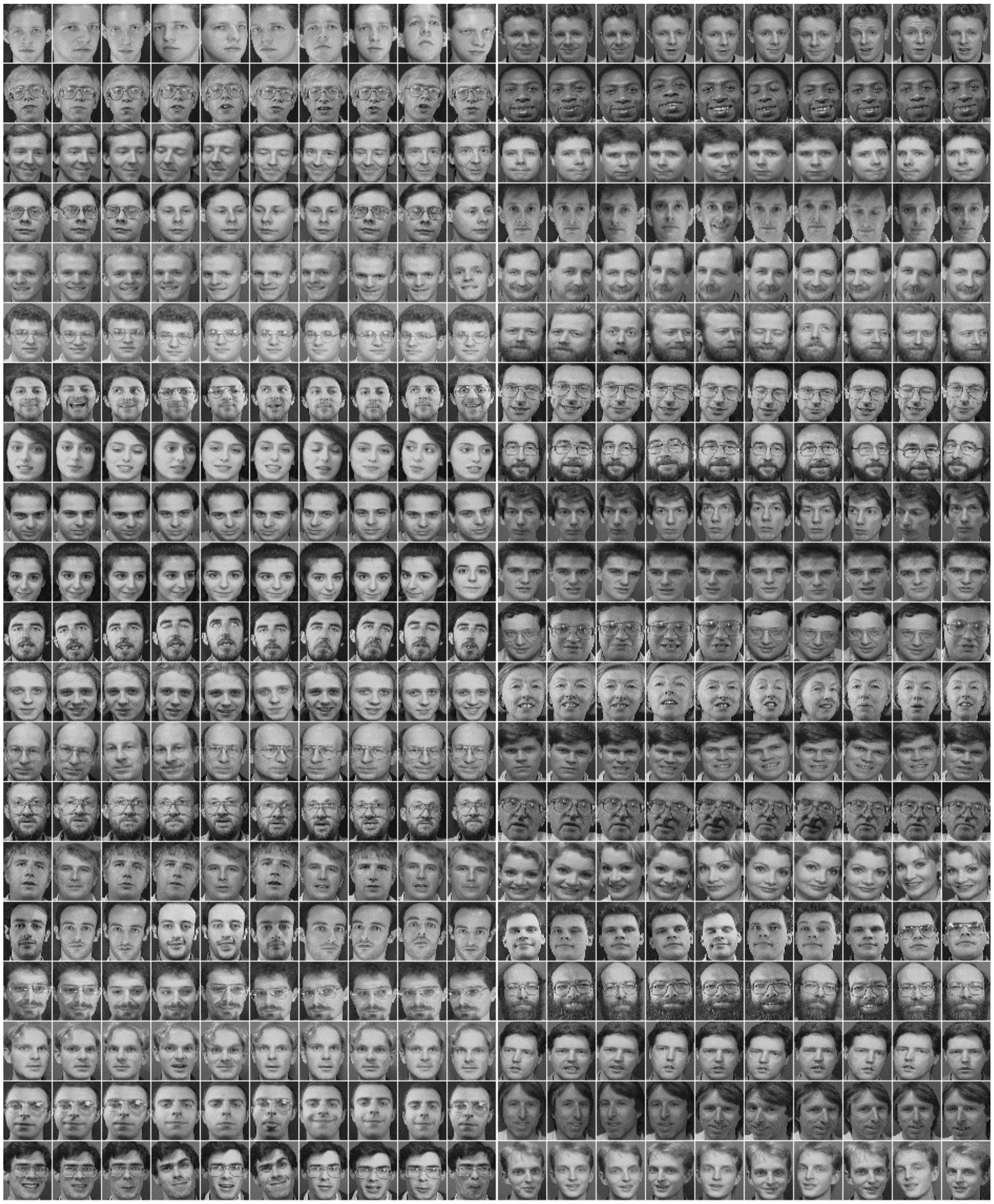


Figure 4.3: Whole set of face images of 40 individuals 10 images per person.

4.1.3. Results for The Olivetti and Oracle Research Laboratory (ORL) face database

The Olivetti and Oracle Research Laboratory (ORL) face database is also used in order to test our method in the presence of head pose variations. There are ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying lighting, facial expressions (open / closed eyes, smiling / not smiling), facial details (glasses / no glasses) and head pose (tilting and rotation up to 20 degrees). All the images were taken against a dark homogeneous background. Figure 4.3 shows the whole set of 40 individuals 10 images per person from the ORL database. We took the first images for each 40 individuals as reference and the rest is used for testing purposes. It is observed that the performance of the method decreased slightly due to the orientation in depth of the head. The proposed method achieved 95.25% correct classification with ORL database. In Figure 4.4 erroneously classified faces of ORL database are presented. These results were expected, since locations are important in feature vectors comparison. Locations of the facial features (eyes, nose, mouth….) are quite changing with rotation of the head.

Recognition performances on ORL database of well-known methods are tabulated in Table 4.2. Although, the reported recognition rates are better for convolutional neural network and line-based methods, it must be noted that these two are using more than one facial image for each individual in training (Figure 4.5). The proposed method achieves %95,25 correct recognition rate by using only one reference facial image for each individual.

Method	Recognition Performance (%)
Eigenface [20]	80.0
Elastic graph matching [23]	80.0
Neural network [19]	96.2
Line based [33]	97.7
Proposed face recognition method using GWT	95.25

Table 4.2: Performance results of wellknown algorithms on ORL database.

(a) (b)

Figure 4.4: Examples of misclassified faces of ORL database, a) reference faces for two individuals, b) misclassified test faces.

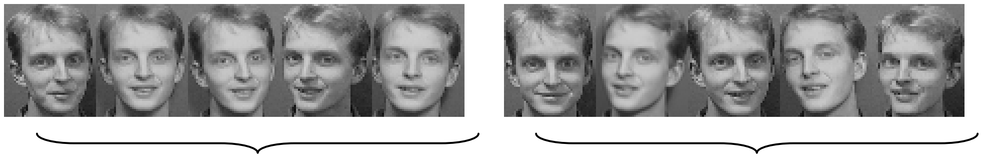


Figure 4.5: Example of different facial images for a person from ORL database that are placed at training or probe sets by neural network and line based algorihms.

4.1.4. Results for FERET

Until recently, there did not exist a common face recognition technology evaluation protocol which includes large databases and standard evaluation methods. The Face Recognition Technology (FERET) program has addressed both issues through the FERET database of facial images and the establishment of the FERET tests. Up to date, 14126 images from 1199 individuals are included in the database.

Primary objectives of FERET test can be stated as:

1. assess the state of the art
2. identify future areas of research
3. measure algorithm performance

The FERET database has made it possible for researchers to develop algorithms on a common database and to report results to the literature based on this database. The results that exist in the literature do not provide a direct comparison between algorithms, since each researcher reports results using different assumptions, scoring methods, and images. The independently administered FERET test allows for a direct quantitative assessment of the relative strengths and weaknesses of different approaches.

The testing protocol is based on a set of design principals. Stating the design principle allows one to assess how appropriate the FERET test is for a particular face recognition algorithm.

There are two sets of images, gallery and probe. The *gallery* is the set of known individuals. An image of an unknown face presented to the algorithm is called a *probe*, and the collection of probes is called the *probe set*.

The main test design principals are:

1. Algorithms can not be trained during testing,
2. Each facial image is treated as a unique face,
3. The similarity score between probe and a gallery image is a function of only those two images.

With the above principles, a similarity between each probe and each gallery image must be generated for performance evaluation.

Although, a time limit for a complete test is assigned as three days while using less than 10 UNIX workstations, time or number of workstations is not recorded for any of the algorithms. Hence, FERET program aims to encourage algorithm development, not code optimization.

The basic models for evaluating the performance of an algorithm are closed and open universes. In a closed universe, every probe is in the gallery. Whereas, in an open universe some probes are not in the gallery. The open universe models verification applications. On the other hand, the closed universe model allows one to ask how good an algorithm is at identifying a probe image; the question is not always “Is the top match is correct?” but “Is the correct answer in the top n matches?”. Such an approach lets one to know how many images have to be examined to get a desired level of performance. The performance statics are reported as cumulative match scores and also the rank is plotted along the horizontal axis, and the vertical axis is the percentage of correct matches. The performance of the algorithm is evaluated on different probe categories.

Note that all the above tests used a single gallery containing 1196 images. The Duplicate I probe images were obtained anywhere between one minute and 1031 days after their respective gallery matches. The harder Duplicate II probe images are a strict subset of the Duplicate I images; they are those taken only at least 18 months after their gallery entries. For assessment of the effect of facial expression, images of subjects with alternate facial expressions (fafb) have been used. There is usually only a few seconds between the capture of the gallery-probe pairs. Finally, some images are obtained under different illumination conditions and gathered under another set (fafc). All these sets are tabulated in Table 4.3. Table 4.4 shows different simulations performed using these sets.

All of the latest results of FERET test is presented in Table 4.5; Cumulative match scores for each of 4 probe sets (dupI, dupII, fafb, fafc) of 14 algorithms that attend to the FERET test and proposed face recognition method are presented. On frontal images taken the same day, typical first choice recognition performance is 95% accuracy. For images taken with a different camera and lighting, typical performance drops to 80% accuracy for the first choice recognition. For images taken one year later, the typical accuracy is approximately 50% (Table 4.5). Note that, even 50% accuracy is 600 times better than a random selection from 1196 faces.

Study	**Goal**
Dup1	Duplicate I
Dup2	Duplicate II
fafb	expression
fafc	illumination

Table 4.3: Probe sets and their goal of evaluation.

Evaluation Task	**Recognized Names**	**Number of faces**	
		Gallery	**Probe Set**
Aging of subjects	Duplicate I	1196	722
Aging of subjects	Duplicate II	1196	234
Facial Expression	fafb	1196	1195
Illumination	fafc	1196	194

Table 4.4: Probe sets for FERET performance evaluation.

Arl_cor is a normalized correlation based algorithm [127]. For normalized correlation, the images were (1) translated, rotated, and scaled so that the center of the eyes were placed on specific pixels and (2) faces were masked to remove background and hair. Arl_ef is a principal components analysis (PCA) based algorithm [20]. These algorithms are developed by U.S. Army Research Laboratory and they provide a performance baseline. In the implementation of the PCA-based algorithm, all images were (1) translated, rotated, and scaled so that the center of the eyes were placed on specific pixels, (2) faces were masked to remove background and hair, and (3) the non-masked facial pixels were processed by a histogram equalization algorithm. The training set consisted of 500 faces. Faces were represented by their projection onto the first 200 eigenvectors and were identified by a nearest neighbor classifier using the L1 metric. ef_hist_dev_ang, ef_hist_dev_anm, ef_hist_dev_l1, ef_hist_dev_l2, ef_hist_dev_md, ef_hist_dev_ml1, ef_hist_dev_ml2 are seven eigenface based system from National Institute of Standards and Technology (NIST) with a common image processing front end and eigenface representation but differing in the distance metric. Algorithms are also tested from Excalibur Corp. (Carlsbad, CA), and from University of Maryland (umd_mar_97) [129, 130]. There are two algorithms developed by MIT Media Laboratory using a version of eigenface transform; mit_mar_95 is the algorithm is the same algorithm that was tested in March 1995 [86], algorithm retested in order to measure improvements, and mit_sep_96 is the algorithm developed since March 1995 [22]. And finally the usc_mar_97 is an elastic graph matching algorithm [23] from Southern California University.

	First Choice Recognition Ratio			
	dupI	dupII	fafb	fafc
arl_cor	0.363	0.171	0.827	0.052
arl_ef	0.410	0.222	0.797	0.186
ef_hist_dev_ang	0.341	0.124	0.701	0.072
ef_hist_dev_anm	0.446	0.209	0.774	0.237
ef_hist_dev_l1	0.350	0.132	0.772	0.258
ef_hist_dev_l2	0.331	0.137	0.716	0.041
ef_hist_dev_md	0.422	0.167	0.741	0.232
ef_hist_dev_ml1	0.305	0.128	0.733	0.392
ef_hist_dev_ml2	0.346	0.128	0.772	0.309
excalibur	0.414	0.197	0.794	0.216
mit_mar_95	0.338	0.171	0.834	0.155
mit_sep_96	0.576	0.342	0.948	0.320
umd_mar_97	0.472	0.209	0.962	0.588
usc_mar_97	0.591	0.521	0.950	0.820
Proposed face recognition method using GWT	0.448	0.239	0.963	0.676

Table 4.5: FERET performance evaluation results for various face recognition algorithms.

In Figures 4.6 to 4.9, recognition performances of various face recognition methods are presented. There can be also seen the improvements on algorithms' performances from September 1996 to March 1997. Moreover, In Figure 4.10 average and in Figure 4.11 current upper bound identification performances of above methods on each probe set are presented.

Simulation results show that, face recognition performance of the proposed method is competitive to those of other popular methods, such as elastic graph matching, eigenfaces, etc. Moreover, on fafb and fafc probe sets proposed method achieves higher performance results than the most of the FERET test contenders (Table 4.5). In Figure 4.11-4.15 cumulative performance results of proposed method on each probe set are presented.

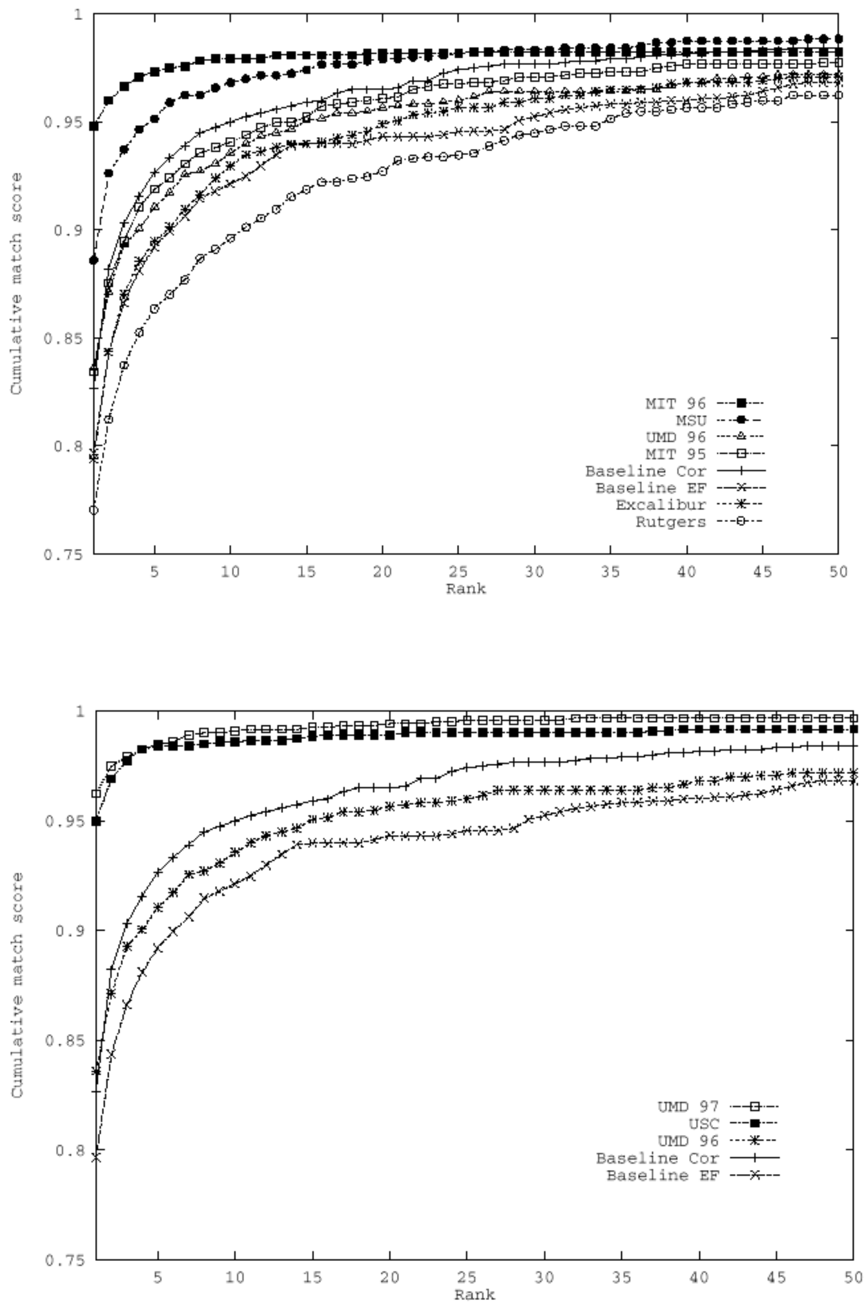
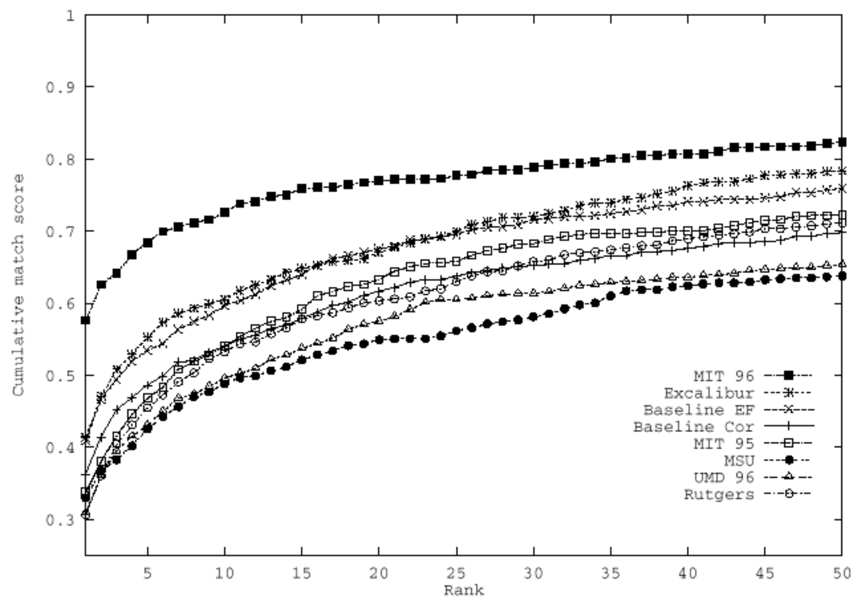


Figure 4.6: Identification performance against fb probes. (a) algorithms tested in September 1996. (b) algorithms tested in March 1997.



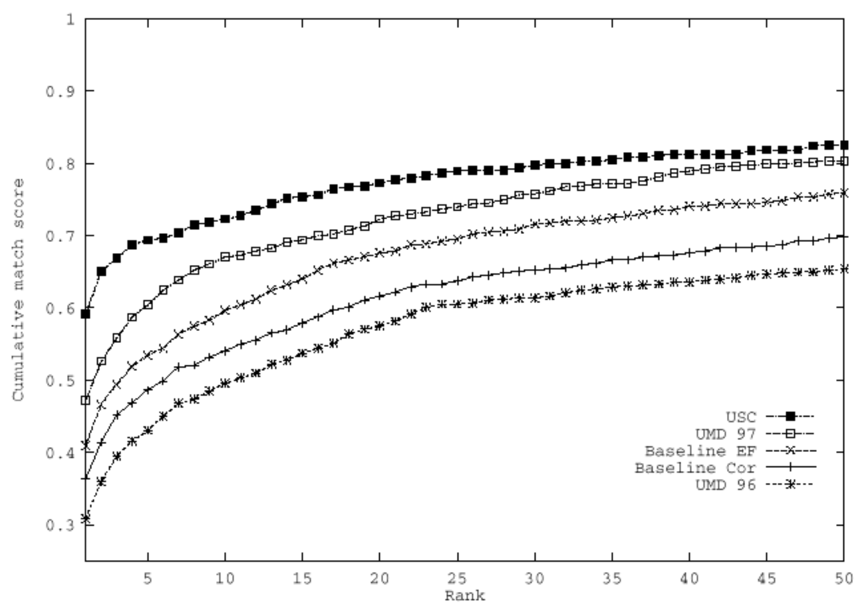


Figure 4.7: Identification performance against duplicate I probes. (a) algorithms tested in September 1996. (b) algorithms tested in March 1997.

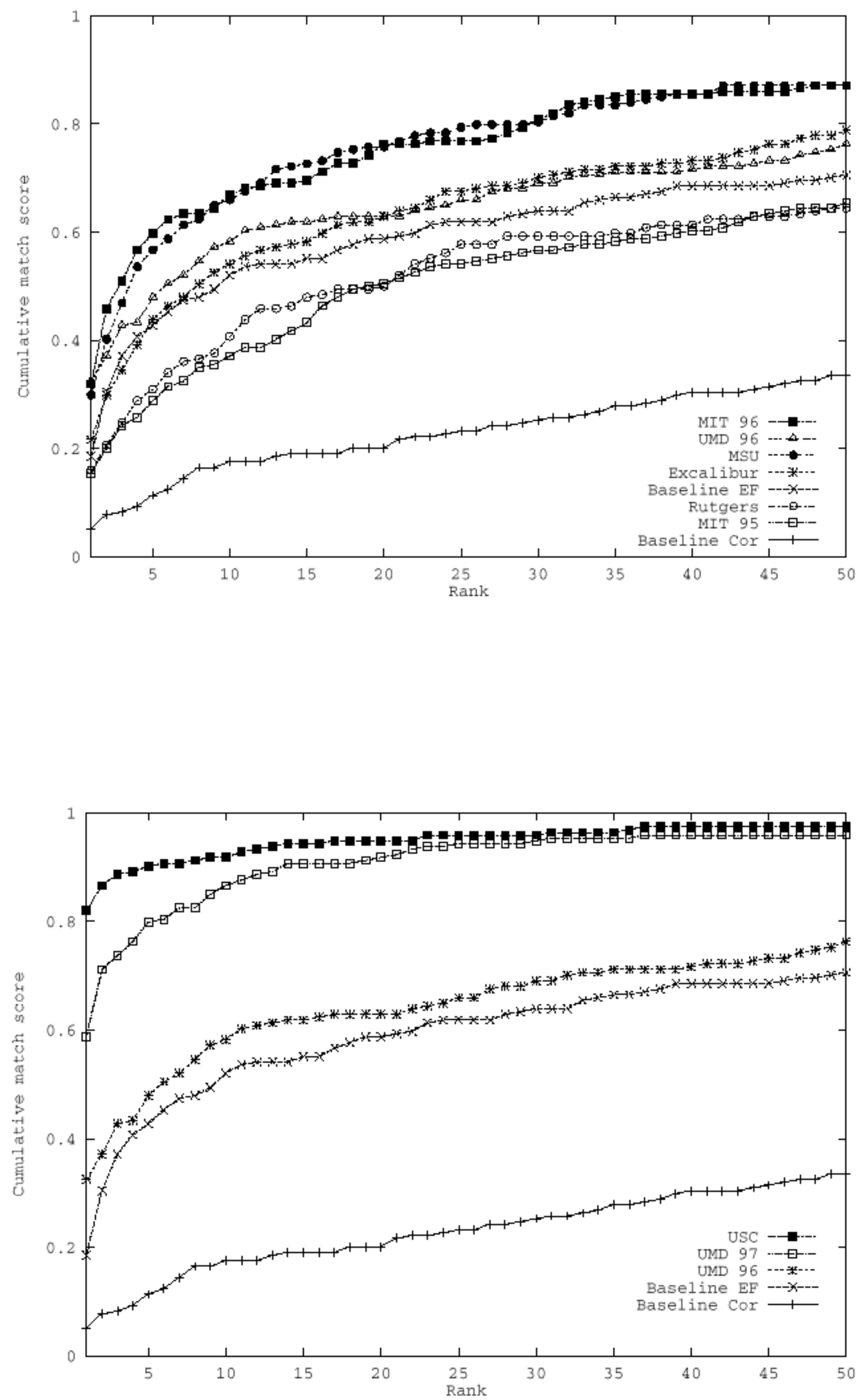
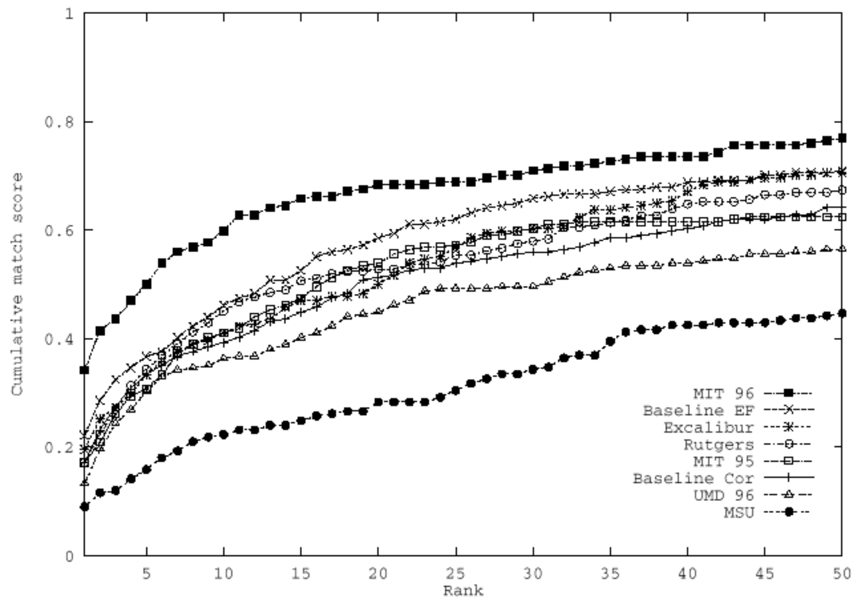


Figure 4.8: Identification performance against fc probes. (a) algorithms tested in September 1996. (b) algorithms tested in March 1997.



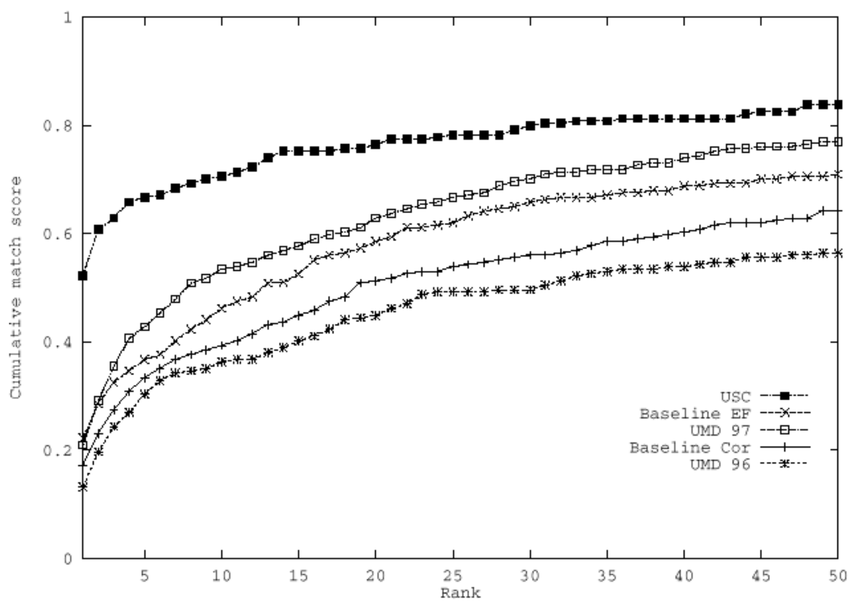


Figure 4.9: Identification performance against duplicate II probes. (a) algorithms tested in September 1996. (b) algorithms tested in March 1997.

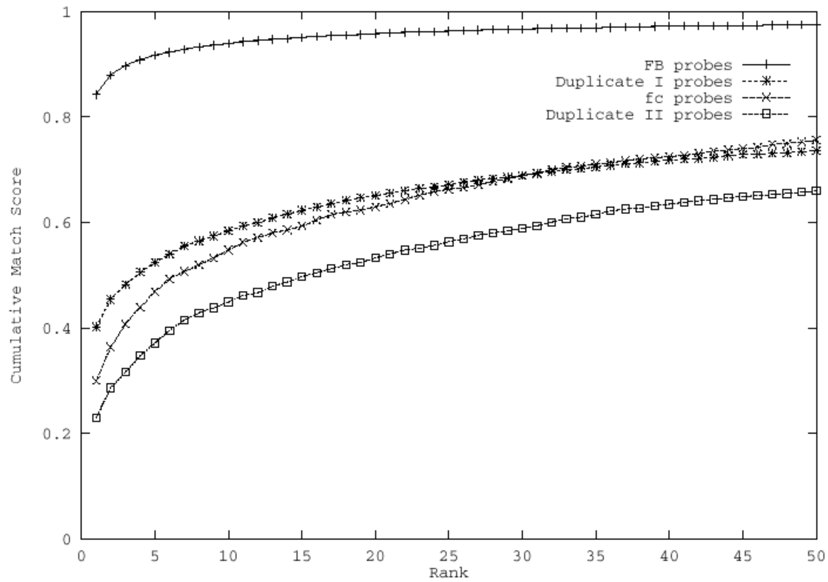


Figure 4.10: Average identification performance of FERET test contenders on each probe category.

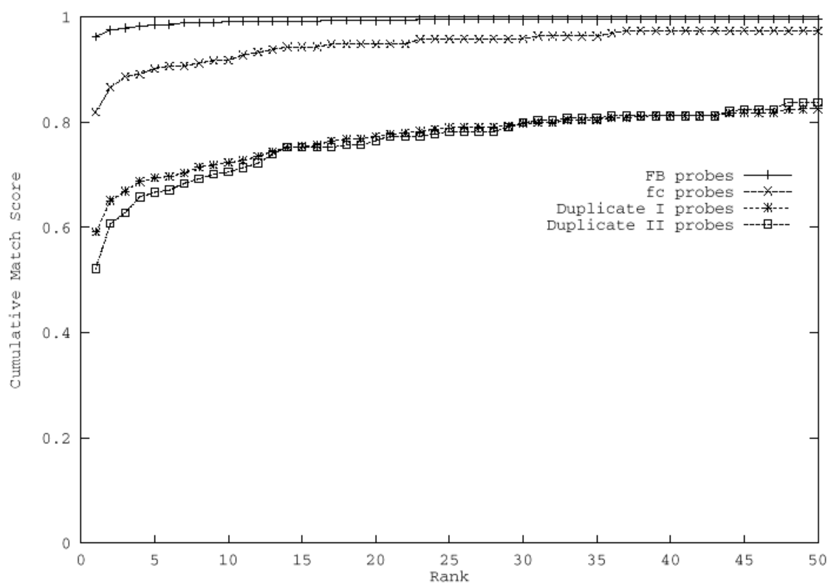


Figure 4.11: Current upper bound identification performance of FERET contender for each probe category.

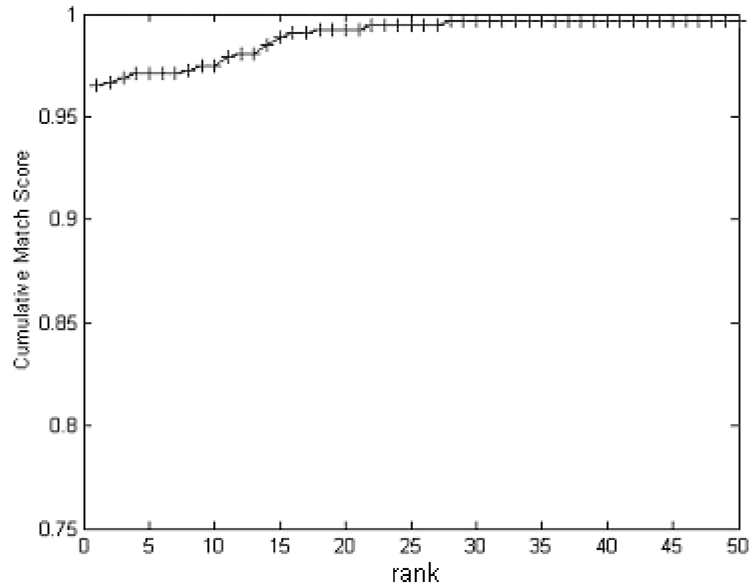


Figure 4.12: Identification performance of proposed method against fb probes.

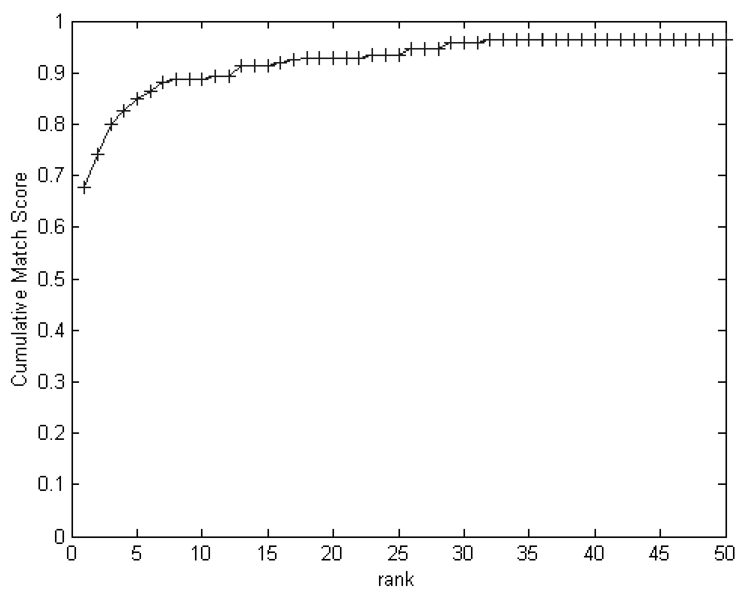


Figure 4.13: Identification performance of proposed method against fc probes.

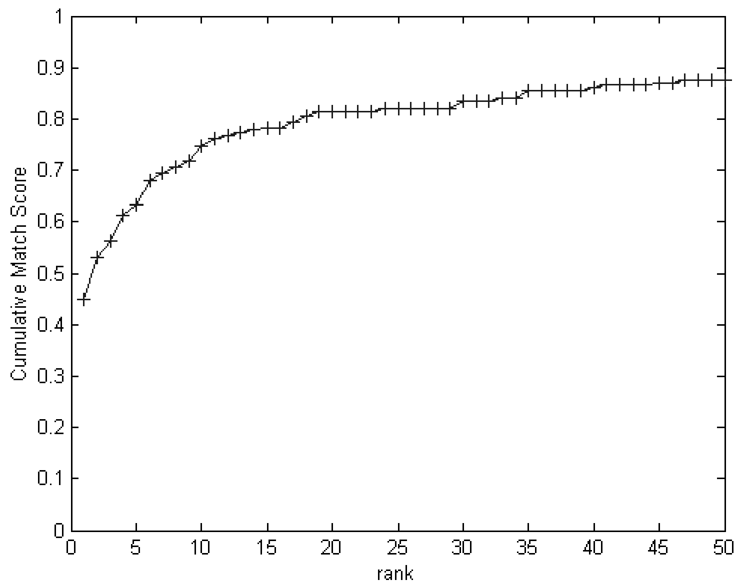


Figure 4.14: Identification performance of proposed method against duplicate I probes.

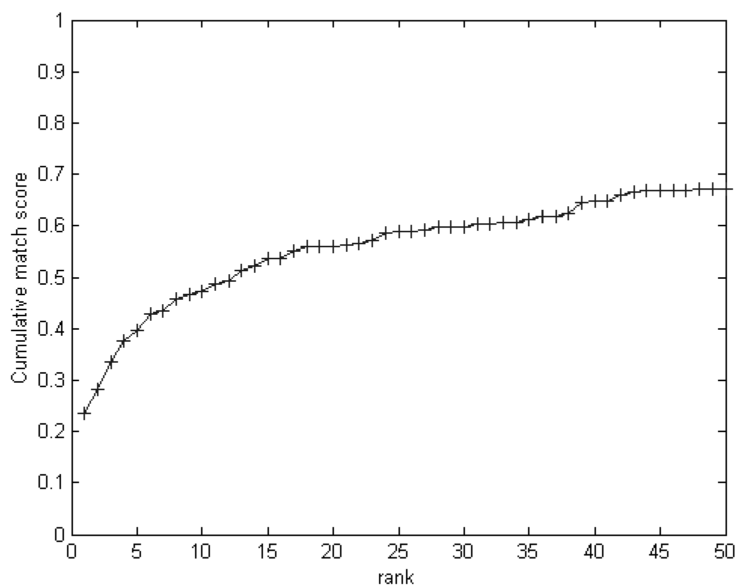


Figure 4.15: Identification performance of proposed method against duplicate II probes.

CHAPTER 5

CONCLUSION AND FUTURE WORK

Face recognition has been an attractive field of research for both neuroscientists and computer vision scientists. Humans are able to identify reliably a large number of faces and neuroscientists are interested in understanding the perceptual and cognitive mechanisms at the base of the face recognition process. Those researches illuminate computer vision scientists' studies. Although designers of face recognition algorithms and systems are aware of relevant psychophysics and neurophysiological studies, they also should be prudent in using only those that are applicable or relevant from a practical/implementation point of view.

Since 1888, many algorithms have been proposed as a solution to automatic face recognition. Although none of them could reach the human recognition performance, currently two biologically inspired methods, namely eigenfaces and elastic graph matching methods, have reached relatively high recognition rates.

Eigenfaces algorithm has some shortcomings due to the use of image pixel gray values. As a result system becomes sensitive to illumination changes, scaling, etc. and needs a beforehand pre-processing step. Satisfactory recognition performances could be reached by successfully aligned face images. When a new face attend to the database system needs to run from the beginning, unless a universal database exists.

Unlike the eigenfaces method, elastic graph matching method is more robust to illumination changes, since Gabor wavelet transform of images is being used, instead of directly using pixel gray values. Although recognition performance of elastic graph matching method is reported higher than the eigenfaces method [83], due to its computational complexity and execution time, the elastic graph matching approach is less attractive for commercial systems. Although using 2-D Gabor wavelet transform seems to be well suited to the problem, graph matching makes algorithm bulky. Moreover, as the local information is extracted from the nodes of a predefined graph, some details on a face, which are the special characteristics of that face and could be very useful in recognition task, might be lost.

In this thesis, a new approach to face recognition with Gabor wavelets is presented. The method uses Gabor wavelet transform for both finding feature points and extracting feature vectors. From the experimental results, it is seen that proposed method achieves better results compared to the graph matching and eigenface methods, which are known to be the most successive algorithms.

Although the proposed method shows some resemblance to graph matching algorithm, in our approach, the location of feature points also contains information about the face. Feature points are obtained from the special characteristics of each individual face automatically, instead of fitting a graph that is constructed from the general face idea. In the proposed algorithm, since the facial features are compared locally, instead of using a general structure, it allows us to make a decision from the parts of the face. For example, when there are sunglasses, the algorithm compares faces in terms of mouth, nose and any other features rather than eyes. Moreover, having a simple matching procedure and low computational cost proposed method is faster than elastic graph matching methods. Proposed method is also robust to illumination changes as a property of Gabor wavelets, which is the main problem with the eigenface approaches. There is no training as in many supervised approaches, such as neural networks. A new facial image can also be simply added by attaching new feature vectors to reference gallery while such an operation might be quite time consuming for systems that need training.

The algorithm proposed by Manjunath et. al. [36] that shows some similarity to our algorithm, especially in terms of the utilized features. However, Manjunath et. al. Disregard a point while using an additional topology cost. Their topology cost is defined as the ratio of the vectoral distances of feature point pairs between two face images. During simulations, it is observed that the locations of feature points, found from Gabor responses of the face image, can give small deviations between different conditions (expression, illumination, having glasses or not, rotation, etc.), for the same individual. Therefore, an exact measurement of corresponding distances is not possible unlike the geometrical feature based methods. Moreover, due to automatical feature detection, features represented by those points are not explicitly known, whether they belong to an eye or a mouth, etc. Giving an information about the match of the overall facial structure, the locations of feature points are very important. However using such a topology cost amplifies the small deviations of the locations of feature points that are not a measure of match.

Gabor wavelet transform of a face image takes 1.5 seconds, feature extraction step of a single face image takes 0.3 seconds and matching an input image with a single gallery image takes 0.15 seconds on a Pentium III 550 Mhz PC. Note that above execution times are measured without code optimization.

Although recognition performance of the proposed method is satisfactory by any means, it can further be improved with some small modifications and/or additional pre-processing of face images. Such improvements can be summarized as;

- Since feature points are found from the responses of image to Gabor filters separately, a set of weights can be assigned to these feature points by counting the total times of a feature point occurs at those responses.
- A motion estimation stage using feature points followed by an affine transformation could be applied to minimize rotation effects. This process will

not create much computational complexity since we already have feature vectors for recognition. By the help of this step face images would be aligned.

- When there is a video sequence as the input to the system, a frame giving the "most frontal" pose of a person should be selected to increase the performance of face recognition algorithm. This could be realized by examining the distances between the main facial features which can be determined as the locations that the feature points become dense. While trying to maximize those distances, for example distance between two eyes, existing frame that has the closest pose to the frontal will be found. Although there is still only one frontal face per each individual in the gallery, information provided by a video sequence that includes the face to be recognized would be efficiently used by this step.
- As it is mentioned in problem definition, a face detection algorithm is supposed to be done beforehand. A robust and successive face detection step will increase the recognition performance. Implementing such a face detection method is an important future work for successful applications.
- In order to further speed up the algorithm, number of Gabor filters could be decreased with an acceptable level of decrease in recognition performance.

It must be noted that performance of recognition systems is highly application dependent and suggestions for improvements on the proposed algorithm must be directed to a specific purpose of the face recognition application.

REFERENCES

[1] V. Bruce, *Recognizing Faces*. London: Erlbaum, 1988.

[2] G. Davies, H. Ellis, and E. J. Shepherd, *Perceiving and Remembering Faces*, New York: Academic, 1981.

[3] H. Ellis, M. Jeeves, F. Newcombe, and A. Young, *Aspects of Face Processing*. Dordrecht: Nijhoff, 1986.

[4] R. Baron, "Mechanisms of human facial recognition," *Int. J. Man-Machine Studies*, vol. 15, pp. 137-178, 1981.

[5] D. C. Hay and A. W. Young, "The human face," *Normality and Pathology in Cognitive function,* A. W. Ellis Ed. London: Academic, 1982, pp. 173-202.

[6] S. Carey, "A case study: Face Recognition," *Explorations in the Biological Language*, E. Walker Ed. New York: Bradford, 1987, pp. 175-201.

[7] S. Carey, R. Diamond, and B. Woods, "The development of face recognition- A maturational component?" *Develop. Psych.*, vol. 16, pp. 257-269, 1980.

[8] A. P. Ginsburg, "Visual Information processing based on spatial filters constrained by biological data," *AMRL tech. Rep.*, pp. 78-129, 1978.

[9] A. G. Goldstein, "Facial feature variation: Anthropometric Data II," *Bull. Psychonomic Soc.*, vol.13, pp. 191-193, 1979.

[10] A. G. Goldstein, "Face related variation of facial features: Anthropometric Data I," *Bull. Psychonomic Soc.*, vol.13, pp. 187-190, 1979.

[11] L. D. Harmon, “The recognition of faces.” *Scientific American*, vol. 229, pp. 71-82, 1973.

[12] D. Perkins, “ A definition of caricature and recognition,” *Studies in the Anthropology of Visual Commun*., vol. 2, pp. 1-24, 1975.

[13] J. Sergent, “Microgenesis of face perception,” *Aspects of Face Processing*, H. D. Ellis, M. A. Jeeves, F. Newcombe, and A. Young Eds. Dordrecht: Nijhoff, 1986.

[14] T. Kanade, “Picture processing by computer complex and recognition of human faces”. *Technical report*, Kyoto University, Dept. of Information Science, 1973.

[15] S. Lin, S. Kung, and L. Lin, “Face Recognition / Detection by Probabilistic Decision-Based Neural Network,” *IEEE Trans. Neural Networks*, vol.8, pp.114-132, 1997.

[16] R. Brunelli, T. Poggio, “Face Recognition: Features vs. Templates,” *IEEE Trans. on PAMI*, Vol. 12, No. 1, Jan. 1990.

[17] M. H. Yang, N. Ahuja, and D. Kriegman, “A survey on face detection methods,” *IEEE Trans. On Pattern analysis and Machine Intelligance*, to appear 2001.

[18] S. Ranganath and K. Arun, “Face Recognition Using Transform Features and Neural Network,” *Pattern Recognition*, vol. 30, pp. 1615-1622, 1997.

[19] S. Lawrence, C. Giles, A. Tsoi, and A. Back, “Face Recognition: A Convolutional Neural Network Approach,” *IEEE Trans. on Neural Networks*, vol. 8, pp. 98-113, 1997.

[20] M. Turk and A. Pentland. “Eigenfaces for recognition.” *Journal of Cognitive Science*, pp.71-86, 1991.

[21] P. Belhumeur, J. Hespanha, and D. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class Specific linear projection,” *IEEE Trans. on PAMI*, vol.19, no.7, 1997.

[22] B. Moghaddam, C. Nastar, and A. Pentland, “Bayesian Face Recognition using Deformable Intensity Surfaces,” *IEEE Conference on CVPR*, San Francisco, CA, June 1996.

[23] L. Wiskott, J. M. Fellous, N. Krüger and Christoph von der Malsburg, “Face Recognition by Elastic Graph Matching,” *In Intelligent Biometric*

Techniques in fingerprint and Face Recognition, CRC Press, Chapter 11, pp. 355-396, 1999.

[24] J. G. Daugman, "Complete Discrete 2-D Gabor Transform by Neural Networks for Image Analysis and Compression," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 36, no.7, pp.1169-1179, 1988.

[25] M. J. Lyons, J.Budynek and S. Akamatsu, "Automatic Classification of Single Facial Images," *IEEE Trans. on PAMI*, vol. 21, no.12, December 1999.

[26] N. Krüger, M. Pötzsch and C. von der Malsburg, "Determining of Face Position and Pose with a learned Representation Based on Labeled Graphs," *Image and Vision Computing*, vol. 15, pp.665-673, 1997.

[27] L. Wiskott, J. M. Fellous, N. Krüger and Christoph von der Malsburg, "Face Recognition and Gender Determination," *Int'l Workshop on Automatic Face Recognition and Gesture Recognition*, pp.92-97, 1995.

[28] Stirling database, http://pics.psych.stir.ac.uk.

[29] Purdue face database, http://www.purdue.edu.

[30] Olivetti & Oracle Research Laboratory, The Olivetti & Oracle Research Laboratory Face Database of Faces, http://www.cam-orl.co.uk/facedatabase.html

[31] A. Yılmaz, M. Gökmen, "Eigenhill vs. Eigenface and Eigenedge,"Proc. of International Conf. on Pattern Recognition, Barcelona Spain, 2000.

[32] J. Zhang, Y. Yan, and M. Lades, "Face Recognition: Eigenface, Elastic Matching, and Neural Nets," *Proc. IEEE*, vol. 85, pp.1423-1435, 1997.

[33] O. de Vel and S. Aeberhard, "Line-Based Face Recognition Under Varying Pose," *IEEE Trans. on PAMI*, vol. 21, no. 10, October 1999.

[34] B. Moghaddam, W. Wahid, and A. Pentland, "Beyond Eigenfaces: Probabilistic Matching for Face Recognition," *The 3rd IEEE Int'l Conference on Automatic Face and Gesture Recognition*, Nara, Japan, April 1998.

[35] B. Duc, S. Fisher, and J. Bigün, "Face Authentication with Gabor Information on Deformable Graphs," *IEEE Trans. On Image Proc.*, vol.8, no.4, pp.504-515, 1999.

[36] B.S. Manjunath, R. Chellappa, and C. von der Malsburg, “A Feature Based Approach to Face Recognition,” *Proc. of International Conf. on Computer Vision*, 1992.

[37] N.M. Allinson, A.W. Ellis, B.M. Flude, and A.J. Luckman, “A connectionist model of familiar face recognition,” *IEE Colloquium on Machine Storage and Recognition of Faces*, vol. 5, pp. 1-10 (Digest No: 1992/017), 1992.

[38] R.J. Baron, “Mechanisms of human facial recognition,” *International Journal on Machine Studies*, vol. 15, pp. 137-178, 1981.

[39] P. Benson and D. Perrett, “Face to face with the perfect image,” *New Scientist*, pp. 32-35, 22 February 1992.

[40] P.J. Benson and D.I. Perrett, “Perception and recognition of photographic quality facial caricatures: implications for the recognition of natural images,” *Face Recognition*, pp. 105-135, Laurence Erlbaum Associates, 1991.

[41] T. Valentine, V. Bruce, “Recognizing familiar faces: the role of distinctiveness and familiarity”, *Canadian Journal of Psychology*, vol. 40, pp. 300-305, 1986.

[42] V. Bruce and A.W. Young, “Understanding face recognition,” *British Journal of Psychology*, vol. 77, pp. 305-327, 1986.

[43] N. Cantor and W. Mischel, “Prototypes in person perception,” *Advances in Experimental Social Psychology*, vol. 12, pp. 3-52, Academic Press, 1979.

[44] G.W. Cottrell and M. Fleming, “Face recognition using unsupervised feature extraction,” *International Neural Network Conference*, vol. 1, pp. 322-325, 1990.

[45] I. Craw, “Recognizing face features and faces,” *IEE Colloquium on Machine Storage and Recognition of Faces*, vol. 7, pp. 1-4 (Digest No: 1992/017), 1992.

[46] I. Craw and P. Cameron, “Face recognition by computer,” *British Machine Vision Conference*, pp. 488-507, Springer-Verlag, 1992.

[47] J. Daugman, “Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters,” *Journal of the Optical Society of America*, vol. 2(7), 1160-1169, July 1985.

[48] J.G. Daugman, “High confidence visual recognition of persons by a test of statistical independence,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11), 1993.

[49] R. Diamond and S. Carey, “Why faces are and are not special: an effect of expertise,” *Journal of Experimental Psychology*, vol. 115, pp. 107-117, 1986.

[50] G.D. Forney, “The Viterbi Algorithm,” *Proceedings of the IEEE*, vol. 61(3), pp. 268-278, March1973.

[51] R.C. Fowler, “FINGERPRINT: an old touchstone decriminalized,” *IEEE Spectrum*, pp. 26, February 1994.

[52] R. Gallery and T.I.P. Trew, “An architecture for face classification,” *IEE Colloquium on Machine Storage and Recognition of Faces*, vol. 2, pp. 1-5 (Digest No: 1992/017), 1992.

[53] F. Galton, “Personal identification and description 1,” *Nature*, pp. 173-177, 21 June1888.

[54] Sir Francis Galton, “Personal identification and description-II”, *Nature* 201-203, 28 June 1888.

[55] L. Gillick and S.J. Cox, “Some statistical issues in the comparison of speech recognition algorithms,” *Proceedings of the Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 532-535, 1989.

[56] A.J. Goldstein, L.D. Harmon, and A.B. Lesk, “Identification of human faces,” *Proc. of the IEEE*, vol. 59(5), pp. 748-760, 1971.

[57] L.D. Harmon, M.K. Khan, R. Lasch, and P.F. Ramig, “Machine identification of human faces,” *Pattern Recognition*, vol. 13(2), pp. 97-110, 1981.

[58] D.C. Hay and A.W. Young, “The human face,” *Normality and pathology in cognitive functions*, pp. 173-202. Academic Press, 1982.

[59] C.L. Huang and C.W. Chen, Human facial feature extraction for face interpretation and recognition,” *Pattern Recognition*, vol. 25(12), pp.1435-1444, 1992.

[60] T. Hutcheson, “FACE: smile, you're on candid camera,” *IEEE Spectrum*, pp. 28-29, February 1994.

[61] R.A. Hutchinson and W.J. Welsh, “Comparison of neural networks and conventional techniques for feature location in facial images,” *IEE Int.*

Conf. on Artificial Neural Networks, Conf. Publication Number 313, pp. 201-205, 1989.

[62] T. Kanade, “Computer recognition of human faces,” *Interdisciplinary Systems Research*, Birkhauser Verlag, 1977.

[63] M. Kirby and L. Sirovich, “Application of the Karhunen-Loueve procedure for the characterisation of human faces,” *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol.12(1), pp. 103-108, 1990.

[64] T. Kohonen, *Self-organization and associate memory*, Springer-Verlag, 2nd edition, 1988.

[65] A. Lanitis, C.J. Taylor, and T.F. Cootes, “An automatic face identification system using flexible appearance models,” *British Machine Vision Conference*, volume 1, pp. 65-74, BMVA Press, 1994.

[66] L.L. Light, F. Kayra-Stuart, and S. Hollander, “Recognition memory for typical and unusual faces,” *Journal of experimental Psychology: Human Learning and Memory*, vol. 5, pp. 212-228, 1979.

[67] R. Mandelbaum, “SPEECH: just say the word,” *IEEE Spectrum*, page 30, February 1994.

[68] B. Miller, “Vital signs of identity,” *IEEE Spectrum*, pages 22-30, February 1994.

[69] L.R. Rabiner, “A tutorial on Hidden Markov Models and selected applications in speech recognition,” *Proc. of the IEEE*, vol. 77(2), pp. 257-286, 1989.

[70] G. Robertson and I. Craw, “Testing face recognition systems,” *British Machine Vision Conference*, vol. 1, pp. 25-34, BMVA Press, 1993.

[71] A. Samal and P.A. Iyengar, “Automatic recognition and analysis of human faces and facial expressions: A survey,” *Pattern Recognition*, vol. 25(1), pp. 65-77, 1992.

[72] J.W. Shepherd, F. Gibling, and H.D. Ellis, “The effect of distinctiveness, presentation time and delay on face recognition,” *Face Recognition*, pp. 137-145, Laurence Erlbaum Associates, 1991.

[73] D. Sidlauskas, “HAND: give me five,” *IEEE Spectrum*, pp. 24-25, February 1994.

[74] J.E. Siedlarz, “IRIS: more detailed than a fingerprint,” *IEEE Spectrum*, pp. 27, February 1994.

[75] T.J. Stonham, "Practical face recognition and verification with WISARD," A*spects of face processing*, pp. 426-441, Martinus Nijhoff Publishers, 1986.

[76] R. Want and A. Hopper, "Active badges and personal interactive computing objects," *IEEE Trans. On Consumer Electronics*, February 1992.

[77] E. Winograd, "Elaboration and distinctiveness in memory for faces," *Journal of Experimental Psychology: Human Learning and Memory*, vol. 7, pp. 181-190, 1981.

[78] K.H. Wong, H.H.M. Law, and P.W.M. Tsang, "A system for recognizing human faces," *Proc. of the Int. Conf. on Acoustics*, pp. 1638-1642, 1989.

[79] R.K. Yin, "Face recognition by brain-injured patients: a dissociable ability?," *Neuropsychologia*, vol. 8, pp. 395-402, 1970.

[80] A.W. Young and V. Bruce, "Perceptual categories and the computation of 'grand-mother'," *Face Recognition*, pp. 5-49. Laurence Erlbaum Associates, 1991.

[81] A.L. Yuille, P.W. Hallinan, and D.S. Cohen, "Feature extraction from faces using deformable templates," *Int. Journal of Computer Vision*, vol. 8(2), pp. 99-111, 1992.

[82] R. Chellappa, C. Wilson, and S, Sirobey, "Human and machine recognition of faces: A survey," *Proceedings of IEEE*, vol. 83, May 1995.

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

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

[85] M. Muarse and S. Nayar, "Visual learning and recognition of 3-d objects from appearance," *Int. Journal of Computer Vision*, vol.14, pp. 5-24, 1995.

[86] A. Pentland, B. Moghadam, T. Starter, and M. Turk, "View based and modular eigenspaces for face recognition," *Proc. on IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 84-91, 1994.

[87] H. F. S. Akamatsu, T. Sasaki and Y. Suenuga, "A robust face identification scheme- KL expansion of an invariant feature space," *SPIE*

Proc.:: Intelligent Robots and Computer Vision X: Algorithms and Technology, vol. 1607, pp. 71-84, 1991.

[88] P. Belhumeur, J. Hespanha, and D. Kriegman, “Eigenfaces vs Fisherfaces: Recognition using class specific linear projection,” *Proc. of Fourth European Conf. on Computer Vision, ECCV’96*, pp. 45-56, April 1996.

[89] K. Etemad and R. Chellapa, “Face recognition using discriminant eigenvectors,” *Proc. of ICASSP*, 1996.

[90] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.

[91] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. New York, Wiley, 1973.

[92] Y. Cheng, K. Liu, J. Yang, Y. Zang, and N. Gu, “Human face recognition method based on the statistical model of small sample size,” *SPIE Proc.:: Intelligent Robots and Computer Vision X: Alg. And Techn.*, vol 1607, pp. 85-95, 1991.

[93] Z. Hong, “Algebraic feature extraction of image for recognition,” *Pattern Recognition*, vol. 24, pp. 211-219, 1991.

[94] Q Tian, “Comparison of statistical pattern recognition algorithms for hybrid processing, ii: eigenvector-based algorithms,” *Journal of the Optical Society of America*, vol. 5, pp. 1670-1672, 1988.

[95] Y. Cheng, K. Liu, J. Yang, and H. Wang, “A robust algebraic method for human face recognition,” *Proc. of 11th Int. Conf. on Pattern Recognition*, pp. 221-224, 1992.

[96] L. Rabiner and B. Huang, *Fundamentals of speech recognition*, Englewood Cliffs, NJ: Prentice Hall, 1993.

[97] E. Levin and R. Pieraccini, “Dynamic planar warping for optical character recognition,”*ICAASP* , pp.149-152, 1992.

[98] O. Agazzi, S. Kuo, E. Levin, and R. Pieraccini, “ Connected and degraded text recognition using planar HMM,” *ICASSP*, vol. 5, pp. 113-116, 1993.

[99] S. Kuo and O. Agazzi, “Keyword spotting in poorly printed documents using pseudo 2-d HMM,” *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 1994.

[100] O. Agazzi and S. Kuo, “Hidden Markov Models based optical character recognition in presence of deterministic transformations,” *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 1994.

[101] F. Samaria, “Face segmentation for identification using Hidden Markov Models,” *British Machine Vision Conference*, 1993.

[102] F. Samaria and F. Fallside, “Face identification and feature extraction using Hidden Markov Models,” *Image Processing: Theory and Applications*, 1993.

[103] F. Samaria and F. Fallside, “Automated face identification using Hidden Markov Models,” *Proc. of the Int. Conference on Advanced Mechatronics*, 1993.

[104] F. Samaria and A. Harter, “Parameterization of stochastic model for human face identification,” *Proc. Of the second IEEE Workshop on Application of Computer Vision*, 1994.

[105] F. Samaria and S. Young, “HMM based architecture for face identification,” *Image and Computer Vision*, vol. 12, October 1994.

[106] D. E. Rumelhart and J. L. McClellan, *Parallel and Distributed Processing*, vol.1, Cambridge, MA: MIT Press, 1986.

[107] G. W. Cottrell and M. Fleming, “Face recognition using unsupervised feature extraction,” *Proc. Int. Neural Network Conf.*, vol. 1, Paris, France, July 9-13, 1990, pp. 322-325.

[108] H. Bourlard and Y. Kamp, “Auto-association by multilayer perceptrons and singular value decomposition,” *Biological Cybern.*, vol. 59, pp. 291-294, 1988.

[109] David DeMers and G. W. Cottrell, “Non-linear dimension reduction,” *Advances in Neural Information Processing Systems 5*, pages 580-587, San Mateo, CA, 1993. Morgan Kauffmann Publishers.

[110] J. Weng, N. Ahuja, and T. S. Huang, “ Learning recognition and segmentation of 3-d objects from 2-d images,” *Proc. of Int. Conf. on Computer Vision*, ICCV 93, pp. 121-128, 1993.

[111] T. Kohonen, “The self-orginizing map,” *Proc. of the IEEE*, vol. 78, pp. 1464-1480, 1990.

[112] T. Kohonen, *Self Orginizing Maps*, Springer-Verlag, Berlin, Germany, 1995.

[113] G. Chow and X. Li, “Towards a system for automatic facial feature detection,”*Pattern Recognition* , vol. 26, no. 12, pp. 1739-1755, 1993.

[114] L. Stringa, “Eyes detection for face recognition,” *Applied Artificial Intelligence*, vol. 7, pp. 365-382, Oct-Dec 1993.

[115] P. Burt, “Smart sensing within a pyramid vision machine,” *Proc. of IEEE*, vol. 76, pp. 1006-1015, August 1988.

[116] D. Beymer, “Face recognition under varying pose,” *Proc. of 23rd Image understanding Workshop*, vol.2, pp. 837-842, 1994.

[117] M. Lades, J Vorbruggen, J, Buhmann, J. Lange, von der Malsburg, and R. Wurtz, “ Distortion invariant object recognition in the dynamic link architecture,”*IEEE Trans. Comput.*, vol. 42, no. 3, pp. 300-311, 1993.

[118] T. S. Lee, “Image representation using 2-d Gabor wavelets,” *IEEE Trans. On Pattern Analysis Pattern Analysis and Machine Intelligence*, vol. 18, no.10, October, 1996.

[119] J. G. Daugman, “Two dimensional spectral analysis of cortical receptive field profile”, *Vision Research*, vol. 20, pp. 847-856, 1980.

[120] D. Gabor, “Theory of communication,” *J. IEE*, vol. 93, pp. 429-459, 1946.

[121] D. H. Hubel and T. N. Wiesel, “Functional architecture of macaque monkey visual cortex,” *Proc. Royal Soc. B (London)*, vol. 198, pp.1-59, 1978.

[122] S. Marcelja, “Mathematical description of the responses of simple cortical cells,”*J. Optical Soc. Am.,* vol. 70, pp. 1297-1300, 1980.

[123] A. M. Burton, and V. Bruce, and I. Craw, “Modeling face recognition,” *Philosophical Trans. Of the Royal Soc. B*, vol. 1, p. 457-480, 1993.

[124] V. Bruce and S. Langton, “The use of pigmentation and shading information in recognizing the sex and identities of faces,” *Perception*, vol. 23, pp.803-822, 1994.

[125] R. Kemp, G. Pike, P. White, and A. Musselman, “A perception and recognition of normal and negative faces- the role of shape from shading and pigmentation cues,” *Perception*, vol. 25, pp. 37-52, 1996.

[126] R. E. Galper, “Recognition of faces in photographic negative”, *Psychonomic Science*, vol. 19, pp. 207-208, 1970.

[127] H. Moon and P. J. Phillips, 'Analysis of PCA-based face recognition algorithms,' *Empirical Evaluation Techniques in Computer Vision*, IEEE Computer Society Press, Los Alamitos, CA, 1998.

[128] FERET latest evaluation results, http://www.itl.nist.gov/iad/humanid/feret/perf/eval.html

[129] W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant analysis of principal components for face recognition," *In 3rd International Conference on Automatic Face and Gesture Recognition*, pp. 336-341, 1998.

[130] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *J. Opt. Soc. Am. A*, vol. 14, pp.1724-1733, August 1997.