Face Recognition based on Wavelet Transform, Principal Component Analysis and Neural Network

A

Thesis report

submitted towards the partial fulfillment of the

requirements of the degree of

Master of Engineering

in

Electronic Instrumentation and Control Engineering

submitted by

Rahul Dev Nigam

Roll No-800851016

Under the supervision of

Sunil Kumar Singla
Assistant Professor, EIED

DEPARTMENT OF ELECTRICAL AND INSTRUMENTATION ENGINEERING

THAPAR UNIVERSITY, PATIALA – 147004

JULY 2010

DECLARATION

*I hereby declare that the report entitled **"Face Recognition based on Wavelet Transform, Principal Component Analysis and Neural Network."** is an authentic record of my own work carried out as requirements for the award of degree of M.E. (Electronic Instrumentation & Control) at Thapar University, Patiala, under the guidance of Mr. Sunil Kumar Singla (Assistant Professor, EIED) during January to July 2010.*

Date: 14 July 2010

(Rahul Dev Nigam)

Roll No. - 800851004

It is certified that the above statement made by the candidate is correct to best of my knowledge and belief.

Sunil Kumar Singla

Assistant Professor, EIED,

Thapar University, Patiala

S. Ghosh 15/7/10

Dr. Smarajit Ghosh

Professor & Head, EIED,

Thapar University, Patiala

Dr. R.K. Sharma

Dean of Academic Affairs

Thapar University, Patiala

ACKNOWLEDGEMENT

Words are often too less to reveals one's deep regards. An understanding of the work like this is never the outcome of the efforts of a single person. I take this opportunity to express my profound sense of gratitude and respect to all those who helped me through the duration of this thesis.

First of all I would like to thank the Supreme Power, one who has always guided me to work on the right path of the life. Without His grace this would never come to be today's reality.

This work would not have been possible without the encouragement and able guidance of my supervisor, Mr. Sunil Kumar Singla. Their enthusiasm and optimism made this experience both rewarding and enjoyable. Most of the novel ideas and solutions found in this thesis are the result of our numerous stimulating discussions. Their feedback and editorial comments were also invaluable for the writing of this thesis.

I would like to express my deep sense of gratitude towards Dr. Smarajit Ghosh, Professor and Head, EIED, Thapar University, Patiala who has been a constant source of inspiration for me throughout this work.

I would also like thank all the faculty members of the department and my friends who directly or indirectly helped me in completion of my thesis.

No words of thanks are enough for my dear parents whose support and care makes me stay on earth. Thanks to be with me.

Date: 14 July 2010

Place: Thapar University, Patiala Rahul Dev Nigam

ABSTRACT

Face is a primary focus of attention in social intercourse, playing a major role in conveying identity and emotion. The dictionary meaning of FACE is “The surface of the front of the head from the top of the forehead to the base of the chin and from ear to ear”. Face recognition records the spatial geometry of distinguishing features of the face. Face recognition has been an interesting issue for both neuroscientists and computer engineers dealing with artificial intelligence (AI). A human can identify face easily, whereas, for a computer to recognize the face, the face area should be detected, features are required to be extracted and comparison with the features already stored in the database is required. A key potential advantage of a machine system is its memory capability, whereas, human face recognition system the important feature is the parallel processing capability.

In this work, new face recognition system based on wavelet transform and principal component analysis using back propagation algorithm neural network has been presented. The HAAR is used to form the coefficients matrix for the detection of the face. The image feature vector is obtained by computing principal component analysis from the coefficient matrix of discrete wavelet transform. The Eigen faces approach is then used to reduce the dimension of the face vectors. Reduced feature vector are used for further classification using neural network. Neural network is used to create the face database and recognize and authenticate the face by using the weights.

The proposed work deals with two problems, namely facial expression and illumination. HAAR wavelet transform enhances the contrast as well as edges of the face images and works efficiently in wide range of illumination changes.

Our experiments have been conducted on the YALE database to obtain the optimum learning rate which comes out to be 0.5 in this case considering the final goal and the success rate. The recognition rate

obtained is 97.92% when 52 components from PCA have been selected and 50 neurons in the hidden layer have been used.

TABLE OF CONTENTS

Contents	**Page No.**

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

The dictionary meaning of FACE is "The surface of the front of the head from the top of the forehead to the base of the chin and from ear to ear" [1].

Face is a primary focus of attention in social intercourse, playing a major role in conveying identity and emotion. The human ability to recognize faces is remarkable. People can recognize thousands of faces learned throughout their lifetime and identify familiar faces at a glance even after years of separation. This skill is quite robust, despite large changes in the visual stimulus due to viewing conditions, expression, aging, and distractions such as glasses, beards or changes in hair style.

The face is a central sense organ complex, for those animals that have one, normally on the ventral surface of the head, and can depending on the definition in the human case, include the hair, forehead, eyebrow, eyelashes, eyes, nose, ears, cheeks, mouth, lips, philtrum, teeth, skin, and chin. The face has uses of expression, appearance, and identity amongst others. It also has different senses like olfaction, taste, hearing, and vision [2].

1.1.1 Individuality and recognition

The face is the feature which best distinguishes a person, and there are "special" regions of the human brain, such as the fusiform face area (FFA), which when damaged prevent the recognition of the faces of even intimate family members. The pattern of specific organs such as the eyes or parts thereof is used in biometric identification to uniquely identify individuals. The fusiform face area (FFA) is a part of the human visual system which might be specialized for facial recognition, although there is some evidence that it also processes categorical information about other objects, particularly familiar ones [3].

The human face's proportions and expressions are important to identify origin, emotional tendencies, health qualities, and some social information. From birth, faces are important in the individual's social interaction. Face perceptions are very complex as the face expressions involve vast involvement of areas in the brain. Sometimes damaged parts of the brain can cause specific impairments in understanding faces or prosopagnosia.

Faces are special because they are involved in a multitude of behavioral/social functions and draw on a combination of specific processes, as well as other more general processes.

Faces are complex visual stimuli, not easily described by simple shapes or patterns; yet people have the ability to recognize familiar faces at a glance after years of separation. Quite often we strain to see the resemblance between a picture (e.g., a driver’s license photo) and the real person, and sometimes we are greeted in a friendly, familiar manner by someone we do not remember ever seeing before. Although face recognition in humans may be impressive, it is far from perfect. Yet there is something about the perception of faces that is very fundamental to the human experience. Early in life we learn to associate faces with pleasure, fulfilment, and security. As we get older, the subtleties of facial expression enhance our explicit communication in myriad ways. The face is our primary focus of attention in social intercourse; this can be observed in interaction among animals as well as between humans and animals. The face, more than any other part of the body, communicates identity, emotion, race, and age, and is also quite useful for judging gender, size, and perhaps even character [4].

1.2 The Visual Face

Faces play a vital role in our daily lives with their varied repertoire of intricate and often subtle functions. Our faces contain sense organs of which our eyes are the most useful in allowing us to sense the dynamic world around us. But what information do our faces convey to others? One way in which we use our faces to communicate is through the production of audible speech. However, the visual face alone conveys a plethora of useful information. Indeed, speech is accompanied by facial motion as a result of which lip-reading is possible. Facial expressions produced by visible deformation of the face provide us with a guide to judging mood, character and intent. We are constantly ‘reading’ one another’s faces and such processes clearly play an important part in social interaction and communication. All these changes in facial appearance due to expression and speech occur on a rather short time-scale. There are of course more enduring visual cues provided by a person’s face and these allow us to estimate such factors as age, gender, ethnic origin and identity. It is this last use, as a visual cue for identification that will form the main thread of this book. We have begun by pointing out the other functions of faces because identification is inevitably performed in the presence of all the variations in facial appearance to which these give rise.

Faces form a unique class of objects. They share a common structure: the facial features are always configured in a similar way. Given their similarity, it is remarkable that we can discriminate between so many different people based upon facial appearance alone. This seems even more remarkable when we consider that we perform recognition in the presence of widely varying viewing conditions. The perception of faces is highly dynamic, both in space and time and with respect to a given context. In human perception and behaviour it is certainly the case that we use more than just static facial appearances in discriminating between different faces. Visual context constrains our expectation: the object sitting behind a desk is more likely to be human than the object on top of the bookcase! Other sources of information such as body gestures and gait also provide useful cues. It is important to realise that above all, the perception of faces is a spatio-temporal, dynamic vision task. Whilst it is true that we can often identify someone from a static photograph, in the real world we usually observe and identify faces in a dynamic setting. Facial appearance alters due to relative motion of the face and the observer. Changes in lighting and the environment also affect appearance. Therefore, face perception involves rather more than the perception of static pictures of faces. In particular, it would seem to require the ability to detect and track faces as they move through cluttered scenes which are themselves often complex and dynamic.

1.3 The Changing Face

The process of identifying a person from facial appearance has to be performed in the presence of many often conflicting factors which alter facial appearance and make the task difficult. It is therefore important to examine the sources of variation in facial appearance more closely. One can consider that there are two types of variation. A face can change its appearance due to either intrinsic or extrinsic factors. Intrinsic variation takes place independently of any observer and is due purely to the physical nature of the face. Extrinsic variation, on the other hand, arises when the face is observed via the interaction of light with the face and the observer. It is the intrinsic variations that one must learn to understand and interpret. This is largely what we mean by the perception of faces. Perceptual tasks have to be performed consistently and robustly under all sorts of changes in external conditions characterised by the extrinsic sources of variation. In general, faces exhibit many degrees of intrinsic variability which are difficult to characterise analytically. Table 1.1 lists some rather obvious perceptual tasks and the corresponding intrinsic sources of variation. These are not independent and the list is by no means exhaustive. If these perceptual tasks are to be performed effectively, intrinsic variability must be analysed and modelled. The visual appearance of a face also varies due to a host of extrinsic factors such as those highlighted in Table 1.2 [5]. Typically, the appearance of a face alters considerably depending on the illumination conditions and in particular due to self-shadowing. The characteristics of the camera (or eye) used to observe the face also affect the resulting image quality. Other objects present in the scene can cause occlusion and cast shadows as well as altering the nature of the incident light. However, one of the most significant sources of variation is pose change. It is worth emphasising that the pose of a face is determined by the relative three-dimensional (3D) position and orientation of the observer. It is, therefore, an extrinsic rather than an intrinsic source of variation because viewing geometry requires the presence of an observer. The main cause of pose change is relative rigid motion between the observer and the subject. A face undergoes rigid motion when it changes its position and orientation in 3D space relative to the observer. However, a face can also undergo no rigid motion when its 3D shape changes due for example to speech or facial expression. This results in intrinsic variation of appearance. Whilst these two types of motion usually occur together, it is more convenient to treat them separately. If perceiving faces implies interpreting the intrinsic variations in a manner which is invariant to other changes, such invariances are often best achieved if the extrinsic variations are modelled so that their effects can be negated. For instance, the perception of identity should ideally be posing invariant and as such, the ability to determine the pose of a face can play an important role in perceiving identity. In this context, rigid motion provides strong visual cues for understanding pose. Whilst no rigid facial motion is also likely to provide useful cues for identification, it clearly plays an even more important role in communication and perception of expressions.

Table 1.1 Intrinsic sources of variation in facial appearance [5].

Source	Possible Tasks
Identity	Classification, known-unknown, verification, full identification
Facial Expression	Inference of emotion or intention
Speech	Lip-Reading
Sex	Deciding whether male or female
Age	Estimating Age

Table 1.2 Extrinsic sources of variation in Facial Appearance [5].

Source	Effects
Viewing Geometry	Pose
Illumination	Shading, colour, self-shadowing, specular highlights

Imaging Process	Resolution, focus, imaging noise, sampling of irradiant energy distribution, perspective effects
Other Objects	Occlusion, shadowing, indirect illumination

1.4 Computing Faces

Over the last quarter of a century, scientists and engineers have endeavoured to build machines capable of automatic face perception. This effort has been multi-disciplinary and has benefited from areas as varied as computer science, cognitive science, mathematics, physics, psychology and neurobiology. Computer-based face perception is becoming increasingly desirable for many applications including human-machine interfaces, multimedia, surveillance, security, teleconferencing, communication, animation, visually mediated interaction and anthropomorphic environments. Consequently, there has been a strong research effort during recent years in the study and development of computational models, algorithms and computer vision systems for automatic face perception [6, 7, 8, 9, 10].

Broadly speaking, in order to recognise faces or indeed any visual objects, one needs to resolve a stimulus equivalence problem. Visual stimuli in the form of images of a particular object or class of objects should have something in common that differentiates them from images of other objects. Such commonalities should exist regardless of most reasonable extrinsic and intrinsic changes. Computationally then, how can we represent and measure characteristics that remain unique to faces under different conditions? In fact, it is rather difficult and computationally unrealistic to consider a single, general solution to the problem of modelling faces undergoing all the intrinsic and extrinsic variations. In this book, we mainly focus on the perception of faces for identification under extrinsic variations such as pose change caused by movement. In other words, we assume that the computations are either approximately invariant to most intrinsic variations such as age and expression, or that intrinsic variation is constrained so as to have little effect on facial appearance. Although there is evidence that the perception of faces involves dedicated neural hardware, face perception has many aspects in common with our perception of moving objects in general.

In order to understand the nature of visual perception, let us briefly introduce some of the processes involved. In order to perceive objects, artificial and biological vision systems must solve two general problems: that of *segmentation* (also known as *parsing*), and that of *recognition*. The problem of segmentation involves computation to divide images into regions that correspond to bodies of physical objects in the scene. The problem of recognition is to label such bodies as instances of known objects. The segmentation problem is further addressed by two sub-problems known as the problem of *spatial grouping* (also known as *unit formation*) and the *correspondence problem* [11]. Whilst the process of grouping determines which image elements (pixels) belong to a single physical body, correspondence tries to establish associations over time between image elements that are representations of the same scene entity. Both these tasks are non-trivial to accomplish. This is especially true if the perceived objects are constantly in motion and can be partially occluded due to a change of viewpoint: the problem of the *curse of projection* due to the 3D world being under-constrained in its 2D images.

Dynamic perception of faces is necessarily complex. One can readily expect that visual perception of moving objects such as human faces relies upon a range of computational processes including for instance, the measurement of visual cues such as motion and colour, selective attention face detection, pose estimation, view alignment, face tracking, modelling of identity and identification. Such *perceptual processes* are closely coupled since the information extracted and the computation involved in each process is intrinsically dependent upon those of the others. However, in order to understand the computations required when perceiving faces, it is convenient and even necessary to decompose such a process into a number of clearly definable tasks.

1.5 Perceptual Grouping and Focus of Attention

In any visual scene there is always a large amount of information to process. A necessary computational task is *perceptual grouping* which *focuses* *attention* on areas in the field of view where faces are likely to be present. This task is essentially one of determining small attention windows within the visual field where further computation should be directed. Pre-attentive visual cues such as motion and colour are useful for focusing attention. Notice that focus of attention need not involve determining whether faces are present or where exactly faces are located in the scene.

1.5.1 Detection

The function of determining the presence of a face and locating it within the scene is *face detection.* This task requires a model to discriminate faces from all other visual objects or patterns. It does not require any of a face's intrinsic variability to be interpreted. In particular, it does not involve identification. Face detection is also known as *basic-level* or *entry-level* recognition. It can also be considered to involve the tasks of face image segmentation and face alignment.

1.5.2 Tracking

In a dynamic and cluttered visual scene, the location and appearance of a face can change continually. The task of following a face through a visual scene requires *tracking* which in essence involves establishing *temporal correspondence.* Attentional windows and detected face image regions need to be constantly updated, maintaining an appropriate degree of correspondence over time so that points of reference within these windows and regions are consistent.

1.5.3 Identification

Beyond entry-level recognition, the task of *identification* requires a function to discriminate between different faces. In the study of object recognition in general, such a task is often regarded as being *sub-ordinate level* or *within category* recognition where all faces together constitute a category. It might seem that the problem of identification has just been defined. In fact, the exact nature of the identification task can vary quite significantly. Consider, for instance, a database consisting of a set, *y*, of *M* known faces. Several different identification tasks can be envisaged. In fact, at least four tasks can be defined as follows.

1) **Classification:** The task is to identify a face under the assumption that it is a member of *y*.
2) **Known-Unknown:** The task is to decide whether or not a face is a member of *y*.
3) **Verification:** The identity of a face is supplied by some other non-visual means and must be confirmed using face images. This is equivalent to the known-unknown task with $M = 1$.
4) **Full identification:** The task is to determine whether a face is a member of *y* and if so to determine its identity.

It seems clear that any computational treatment of the *dynamic perception of faces* will involve functions that must at least perform the above tasks well. As a result of decomposing face perception in this way we must address another issue in perceptual processing: that of integration.

1.5.4 Perceptual Integration

Despite that the perception of faces, in common with that of other dynamic visual objects, can be conveniently modelled as an assembly of sub-tasks performed independently by a set of functions, such functions can only be effective and even computationally viable if they are closely coupled. This requires the task of *perceptual integration* or *perceptual control.* Closely coupled information processing implies that the performance of any individual process is highly correlated to and dependent upon the effectiveness of the others. In contrast, conventional approaches to vision have often modelled the computations as independent sequential processes. This has been motivated by the need for simplicity and tractability. However, overwhelming neurobiological evidence suggests that perception is only effective if it is performed by closely coupled, co-operative processes with feedbacks [12, 13].

1.5.5 Visual Learning and Adaptation

Humans are born with a certain innate knowledge of faces and subsequently learn to distinguish between different faces. How much knowledge of faces must be hard-wired into the process of face perception and how can this process then bootstrap using this knowledge so as to learn to recognise many different faces? Since visual information is always subject to noise and occlusion due to the generally ill-posed nature of inverse projection, model learning is often more difficult than expected. In fact, learned models need to be updated and tuned to specific recognition and tracking tasks. The ability to adapt previously learned models reflects another aspect of perceptual integration in which learned or hard-wired models are improved during the process of performing a perceptual task. The perception of moving objects and faces can also benefit greatly if recognition not only tracks and matches visual appearance with known models, but also interprets patterns of behaviour. This has been shown to be important in the perception of moving objects in general [14, 15].

1.5.6 Understanding Pose

Among all the extrinsic and intrinsic sources of variation in facial appearance, pose changes are particularly problematic for the tasks of face detection, tracking and identification. One way of coping with pose change is to make pose explicit by estimating it. Although it may seem unnecessary to estimate pose explicitly in order to perform recognition in a manner that is invariant to pose, it will be useful computationally to consider *pose understanding* as a required task.

1.6 Problem Definition and Contribution of Work

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

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

Recognition of faces from a database is a very complex task: lightning condition may vary tremendously; facial expression also varies time to time; face may appear at different orientations.

Central of this work is the fusion of DWT, PCA (Eigen faces) and Neural Network approach for the face recognition. DWT enhances the contrast as well as edges of the face images and works efficiently in wide range of illumination changes. PCA is a basic method used in face feature extraction and detection. The Eigen faces has proven the capability to provide the significant features and reduces the input size for neural network. Thus, Neural Network speed for recognition is raised.

1.7 Thesis Organization

This thesis is organized in the following manner: Chapter 2 deals with the basic concepts of face recognition and Literature Review for face recognition. Major approaches to the face recognition are discussed and a table of comparison is also given at the end of this chapter. Chapter 3 is based on the details of the proposed face recognition method and the actual system developed. Chapter 4 gives the results drawn from the research and finally in Chapter 5, conclusion and possible directions for future work are given.

Chapter 2

Literature Survey

2.1 Introduction

Face recognition is a pattern recognition task performed specifically on faces. It can be described as classifying a face either "known" or "unknown", after comparing it with stored known individuals. It is also desirable to have a system that has the ability of learning to recognize unknown faces.

Face recognition has been an interesting issue for both neuroscientists and computer engineers dealing with artificial intelligence (AI). A healthy human can detect a face easily and identify that face, whereas for a computer to recognize faces, the face area should be detected and recognition comes next. Hence, for a computer to recognize faces, the photographs should be taken in a controlled environment; a uniform background and identical poses makes the problem easy to solve. These face images are called mug shots [16]. From these mug shots, canonical face images can be manually or automatically produced by some preprocessing techniques like cropping, rotating, histogram equalization and masking.

Computational models of face recognition must address several difficult problems. This difficulty arises from the fact that faces must be represented in a way that best utilizes the available face information to distinguish a particular face from all other faces. Faces pose a particularly difficult problem in this respect because all faces are similar to one another in that they contain the same set of features such as eyes, nose, and mouth arranged in roughly the same manner. Face recognition has several advantages over other biometric technologies: It is natural, nonintrusive, and easy to use. Among the six biometric attributes considered by Hietmeyer [17], facial features scored the highest compatibility in a Machine Readable Travel Documents (MRTD) [18] system based on a number of evaluation factors, such as enrollment, renewal, machine requirements, and public perception. Table 2.1 lists some of the applications of face recognition can be used.

Table 2.1 Lists some of the applications of face recognition. [19]

Areas	**Specific Applications**
Entertainment	Video game, virtual reality, training programs
	Human-robot-interaction, human-computer-interaction
Smart cards	Drivers' licenses, entitlement programs
	Immigration, national ID, passports, voter registration
	Welfare fraud
Information security	TV Parental control, personal device logon, desktop logon
	Application security, database security, file encryption
	Intranet security, internet access, medical records
	Secure trading terminals
Law enforcement and surveillance	Advanced video surveillance, CCTV control
	Portal control, postevent analysis
	Shoplifting, suspect tracking and investigation

2.2 Human Face Recognition

The human face recognition system utilizes more than that of the machine recognition system which is just 2-D data. The human face recognition system uses some data obtained from some or all of the senses; visual, auditory, tactile, etc. All these data is used either individually or collectively for storage and remembering of faces. In many cases, the surroundings also play an important role in human face recognition system. It is hard for a machine recognition system to handle so much data and their combinations. However, it is also hard for a human to remember many faces due to storage limitations [20].

The low spatial frequency components are used to clarify the sex information of the individual whereas high frequency components are used to identify the individual. The low frequency components are used for the global description of the individual while the high frequency components are required for finer details needed in the identification process.

Both holistic and feature information are important for the human face recognition system. Studies suggest the possibility of global descriptions serving as a front end for better feature-based perception [20]. If there are dominant features present such as big ears, a small nose, etc. holistic descriptions may not be used. Also, recent studies show that an inverted face (i.e. all the intensity values are subtracted from 255 to obtain the inverse image in the grey scale) is much harder to recognize than a normal face [21].

For humans, photographic negatives of faces are difficult to recognize. But, there is not much study on why it is difficult to recognize negative images of human faces. Also, a study on the direction of illumination [22] showed the importance of top lighting; it is easier for humans to recognize faces illuminated from top to bottom than the faces illuminated from bottom to top.

According to the neurophysicists, the analysis of facial expressions is done in parallel to face recognition in human face recognition system. Some prosopagnosic patients, who have difficulties in identifying familiar faces, seem to recognize facial expressions due to emotions. Patients who suffer from organic brain syndrome do poorly at expression analysis but perform face recognition quite well [16].

2.3 Face Recognition Processing

The most popular approaches to face recognition [23] are based on either (i) the location and shape of facial attributes, such as the eyes, eyebrows, nose, lips, and chin and their spatial relationships, or (ii) the overall (global) analysis of the face image that represents a face as a weighted combination of a number of canonical faces. While the authentication performance of the face recognition systems that are commercially available is reasonable [24], they impose a number of restrictions on how the facial images are obtained, often requiring a fixed and simple background with controlled illumination. These systems also have difficulty in matching face images captured from two different views, under different illumination conditions, and at different times. It is questionable whether the face itself, without any contextual information, is a sufficient basis for recognizing a person from a large number of identities with an extremely high level of confidence. In order for a facial recognition system to work well in practice, it should automatically:

i. Detect whether a face is present in the acquired image.
ii. Locate the face if there is one.
iii. Recognize the face from a general viewpoint (i.e., from any pose) under different ambient conditions.

A face recognition system generally consists of four modules as depicted in Fig. 2.1:

i. Detection
ii. Alignment
iii. Feature extraction
iv. Matching

where localization and normalization (face detection and alignment) are processing steps before face recognition (facial feature extraction and matching) is performed.

Face detection segments the face areas from the background. In the case of video, the detected faces may need to be tracked using a *face tracking* component. *Face alignment* is aimed at achieving more accurate localization and at normalizing faces thereby whereas face detection provides coarse estimates of the location and scale of each detected face. Facial components, such as eyes, nose, and mouth and facial outline, are located; based on the location points, the input face image is normalized with respect to geometrical properties, such as size and pose, using geometrical transforms or morphing. The face is usually further normalized with respect to photometrical properties such illumination and gray scale.

After a face is normalized geometrically and photometrical, *feature extraction* is performed to provide effective information that is useful for distinguishing between faces of different persons and stable with respect to the geometrical and photometrical variations. For *face matching*, the extracted feature vector of the input face is matched against those of enrolled faces in the database; it outputs the identity of the face when a match is found with sufficient confidence or indicates an unknown face otherwise. Face recognition results depend highly on features that are extracted to represent the face pattern and classification methods used to distinguish between faces whereas face localization and normalization are the basis for extracting effective features.

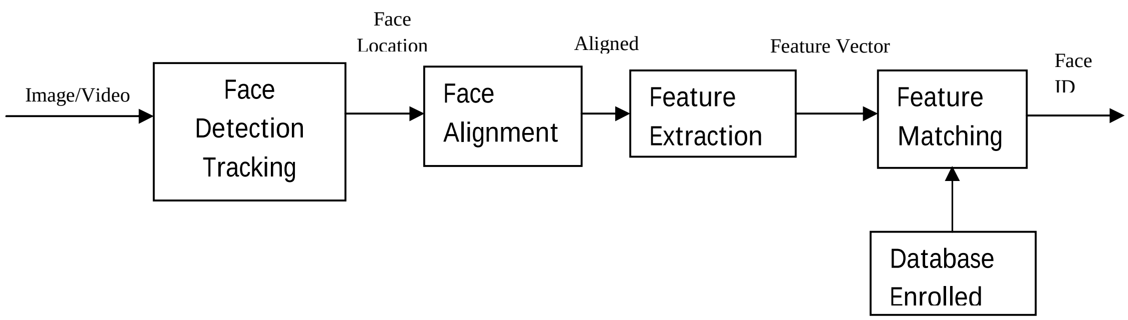


Figure 2.1 Face recognition processing flow [25].

2.4. Machine Recognition of Faces

Although studies on human face recognition were expected to be a reference on machine recognition of faces, research on machine recognition of faces has developed independent of studies on human face recognition. During 1970's, typical pattern classification techniques, which use measurements between features in faces or face profiles, were used [26]. During the 1980's, work on face recognition remained nearly stable. Since the early 1990's, research interest on machine recognition of faces has grown tremendously. The reasons may be;

i. An increase in emphasis on civilian/commercial research projects,

ii. The studies on neural network classifiers with emphasis on real-time computation and adaptation,

iii. The availability of real time hardware,

iv. The growing need for surveillance applications.

Two major approaches are used for machine identification of human faces; geometrical local feature based methods, and holistic template matching based systems. Also, combinations of these two methods, namely hybrid methods, are used. The first approach, the geometrical local feature based one, extracts and measures discrete local features (such as eye, nose, mouth, hair, etc.) for retrieving and identifying faces. Then, standard statistical pattern recognition techniques and/or neural network approaches are employed for matching faces using these measurements [27]. One of the well known geometrical-local feature based methods is the Elastic Bunch Graph Matching (EBGM) technique. The other approach, the holistic one, conceptually related to template matching, attempts to identify faces using global representations [28]. Holistic methods approach the face image as a whole and try to extract features from the whole face region. In this approach, as in the previous approach, the pattern classifiers are applied to classify the image after extracting the features. One of the methods to extract features in a holistic system is applying statistical methods such as Principal Component Analysis (PCA) to the whole image. PCA can also be applied to a face image locally.

Whichever method is used, the most important problem in face recognition is the curse of dimensionality problem. Appropriate methods should be applied to reduce the dimension of the studied space. Working on higher dimension causes over fitting where the system starts to memorize. Also, computational complexity would be an important problem when working on large databases.

2.5 Face Recognition Techniques

Face recognition algorithms can be classified into two broad categories according to feature extraction schemes for face representation: feature-based methods and appearance-based methods [29].

Popular algorithms such as Principal Component Analysis, Linear Discriminant Analysis, Independent Component Analysis, Local Feature Analysis, Correlation Filters, Manifolds and Tensor faces are based on the appearance of the face. Holistic approaches to face recognition have trouble dealing with pose variations. Building image face mosaics like those in [30] [31] have been introduced to deal with the pose variation problem [32].

2.5.1. Eigenfaces (PCA)

Eigenfaces [33] also known as Principal Components Analysis (PCA) find the minimum mean squared error linear subspace that maps from the original N-dimensional data space into an M-dimensional feature space. By doing this, it (where typically M << N) achieves the reduction of dimensionality by covariance matrix corresponding to the large eigenvalues from M eigenvectors. By finding the optimal basis vectors that maximize the total variance of the projected data the resulting basis vectors are obtained (i.e. the set of basis vectors that best describe the data). The optimal basis PCA vectors W are the ones that maximize the following objective function

$$W_{PCA} = \arg\max_{W} \left| W^T S_T W \right| = \left[w_1 w_2 ...... w_m \right] \qquad (2.1)$$

where S_T denotes the total scatter matrix which contains pixel-wise covariances of the face data. Fig. 2.2 shows examples of Eigenfaces generated from the generic training images of Face Recognition Grand Challenge dataset after pre-processing the face images such as normalizing faces for rotation, scale and illumination compensation. PCA not easily generalizes between the classes but it is good for data representation.

Figure 2.2 The first six basis vectors of Eigenfaces [25].

The PCA approach is then used to reduce the dimension of the data by means of data compression basics and reveals the most effective low dimensional structure of facial patterns. This reduction in dimensions removes information that is not useful and precisely decomposes the face structure into orthogonal (uncorrelated) components known as eigenfaces. Each face image may be represented as a weighted sum (feature vector) of the eigenfaces, which are stored in a 1D array. A probe image is compared against a gallery image by measuring the distance between their respective feature vectors. The PCA approach typically requires the full frontal face to be presented each time; otherwise the image results in poor performance. The primary advantage of this technique is that it can reduce the data needed to identify the individual to 1/1000th of the data presented. Fig. 2.3 shows the Standard Eigenfaces: Feature vectors are derived using eigenfaces.

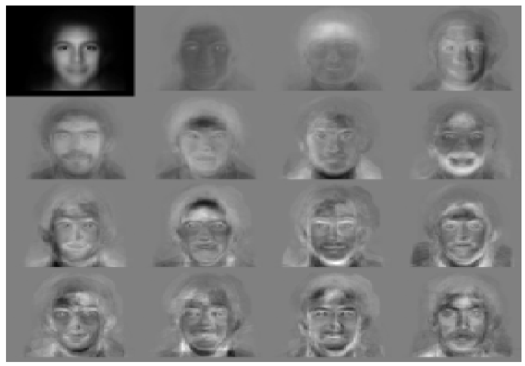


Figure 2.3 Feature vectors are derived using eigenfaces [25].

2.5.2. Linear Discriminant Analysis (LDA) and Fisher faces

LDA is a statistical approach for classifying samples of unknown classes based on training samples with known classes.(Fig 2.4) This technique aims to maximize inter-class variance and minimize intra-class variance. In Fig. 2.4 where each block represents a class, there are large variances between classes, but little variance within classes. When dealing with high dimensional face data, this technique suffers small sample size problem that arises where there are a small number of available training samples compared to the dimensionality of the sample space.

Figure 2.4 Example of six classes using LDA [25].

Linear Discriminant Analysis (LDA) [34] is more suited for discriminate different classes by the projections. This is done by maximizing the ratio of the between-class scatter by the optimal projection vectors (i.e. maximizing class separation in the projected space). The optimal basis vectors of LDA can be denoted as

$$W_{LDA} = \arg\max_{W} \frac{\left|W^T S_B W\right|}{\left|W^T S_W W\right|} \tag{2.2}$$

Where S_B and S_W indicate between-class scatter matrix and within-class scatter matrix respectively.

Typically when dealing with face images (and most other image based pattern recognition problems) the number of training images is smaller than the number of pixels (or equivalently dimensionality of the data), thus the within-class scatter matrix S_W is singular causing problems for LDA. To address this issue [35] first performs PCA to reduce the dimensionality of the data in order to overcome this singular-matrix problem and then applies LDA in this lower-dimensional PCA subspace. Improvement in recognition results was shown using this approach over traditional PCA. The projection vectors from Fisher faces are those that maximize the following objective function:

$$W_{Fisher} = \arg\max_{W} \frac{\left|W^T W_{PCA}^T S_B W_{PCA} W\right|}{\left|W^T W_{PCA}^T S_W W_{PCA} W\right|} \tag{2.3}$$

2.5.3. LDA variants

Direct LDA (DLDA) [36] performs on the eigenvector by the diagonalization technique. Unlike other LDA approaches [37], it immediately diagonalizes between-class scatter matrix and then within-class scatter matrix. The eigenvectors with very small (close to zero) eigenvalues in the S_B can be discarded since they contain no discriminative power, while the eigenvectors with small eigenvalues of the S_W matrix simultaneously being kept, especially the null-space. Another LDA variant is called the Gram-Schmidt LDA (GSLDA) [35] approach avoids computing the inverse of the within-class scatter matrix or performing the diagonalization step needed in DLDA. These methods assert that the most discriminating power for LDA may lie in the null-space of the within scatter matrix which maximizes the Fisher's ratio.

2.5.4. Independent Component Analysis (ICA)

Independent component analysis (ICA) is a computational method for separating a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signals. It is a special case of blind source separation.

Independent Components Analysis (ICA) for face recognition has been applied in [38]. ICA is different from PCA as PCA gives transformed features as uncorrelated by an orthogonal basis for face images, while ICA gives non-orthogonal basis so that transformed features are independent. The basis images developed by PCA depend only on second-order statistics. ICA generalizes the concept of PCA to model higher-order statistical relationships. Original motivation for this decomposition comes from the need to separate audio streams into independent sources without prior knowledge of the mixing process.

General definition

The data is represented by the random vector $x = (x_{1,\dots\dots,}x_2)$ and the components as the random vector $s = (s_{1,\dots\dots,}s_2)$. The task is to transform the observed data x, using a linear static transformation W as

$$s = Wx,$$

Into maximally independent components s measured by some function $F(s_{1,\dots\dots,}s_2)$ of independence.

2.5.5. Local Feature Analysis (LFA)

LFA [39] works on eigen-subspace decomposition and constructs points of locally correlated feature detectors. A selection or sparsification step produces a minimally correlated and topographically indexed subset of features that define the subspace of interest. Due to the changes in localized regions of the object it represents the robustness against variability. The features used in the LFA method are less sensitive to illumination changes and are easier for estimating rotations. It is the main component algorithm in FaceIt [40], for commercial face recognition systems. Local feature analysis (LFA), are used for deriving local topographic representations for any class of objects.

The LFA representations are sparse-distributed and, hence, are effectively low-dimensional and retain all the advantages of the compact representations of the PCA. But, unlike the global eigenmodes, they give a description of objects in terms of statistically derived local features and their positions. We illustrate the theory by using it to extract local features for three ensembles: 2D images of faces without background, 3D surfaces of human heads, and finally 2D faces on a background. The resulting local representations have powerful applications in head segmentation and face recognition.

The LFA representations are low dimensional, and like PCA provide a reduced basis set for collective motions, but they are sparsely distributed and spatially localized. This yields a more reliable assignment of essential dynamics modes across different MD time windows. Also, the intrinsic dynamics of local domains is more extensively sampled than that of globally coherent PCA modes.

2.5.6. Neural Networks (NN) and Support Vector Machines (SVM)

As there is computational complexity in the processing of high-dimensions face data, therefore, Neural Network and Support Vector Machines (SVMs), usually used for low dimensional features spaces. Neural network approaches [41] have been widely explored for feature representation and face recognition. The computational burden increases as the number of people for training increases. Fusion of multiple neural networks classifiers improved the overall performance of face recognition [42].

Face recognition is a K-class problem, where K is the number of known individuals; and SVM is a binary classification method. By reformulating the face recognition problem and reinterpreting the output of the SVM classifier they developed a SVM-based face recognition algorithm. They formulated the face recognition problem in difference space, which models dissimilarities between two facial images. In difference space, they formulated the face recognition as a two class problem. The classes are; dissimilarities between faces of the same person and dissimilarities between faces of different people. By modifying the interpretation of the decision surface generated by SVM, they generated a similarity metric between faces, learned from examples of differences between faces.

A face recognition system using hybrid neural and dual eigenspace methods has been proposed in [43]. However, in general it is not known what exactly the neural network has learned or how it will behave, and for good generalization of data training is required for which significant amount of training is done. Support Vector Machines (SVM) [44] [45] have been successfully applied for object recognition, by utilizing the kernel trick which maps data onto higher-dimensional feature spaces. The SVM hyperplane are found which maximizes the margin of separation so that risk of misclassification for training samples is minimized and also to enable for achieving better generalization of unseen data.

Neural Network approaches have been used in face recognition generally in a geometrical local feature based manner, but there are also some methods where neural networks are applied holistically.

1) ***Feature based Back propagation NN*:** Temdee et al. [46] presented a frontal view face recognition method by using fractal codes which are determined by a fractal encoding method from the edge pattern of the face region (covering eyebrows, eyes and nose). In their recognition system, the obtained fractal codes are fed as inputs to a Back

propagation Neural Network for identifying an individual. They tested their system performance on the ORL face database.

2) ***Dynamic Link Architectures (DLA)****:* Lades et al. [47] presented an object recognition system based on Dynamic Link Architectures, which is an extension of the Artificial Neural Networks. The DLA uses correlations in the fine-scale cellular signals to group neurons dynamically into higher order entities. These entities can be used to code high-level objects, such as a 2-D face image. The face images are represented by sparse graphs, whose vertices are labelled by a multiresolution description in terms of local power spectrum, and whose edges are labelled by geometrical distance vectors. Face recognition can be formulated as elastic graph matching, which is performed in this study by stochastic optimization of a matching cost function.

2.5.7. Elastic Bunch Graph Matching (EBGM)

EBGM relies on the concept that real face images have many nonlinear characteristics that are not addressed by the linear analysis methods discussed earlier, such as variations in illumination (outdoor lighting vs. indoor fluorescents), pose (standing straight vs. leaning over) and expression (smile vs. frown).

To represent individual faces EBGM [29] uses image graphs for constructing dynamic link architecture. A face image is represented by an image graph which a geometrical structure is having various nodes connected by edges. The nodes are located at facial landmarks such as the pupils and the corners of the mouth as shown in Fig. 2.5. The corresponding bunch of image graphs of those images represents a set of training images. A set of complex Gabor wavelet coefficients (or Gabor jets) are used as local features at each node. The Gabor jet is a node on the elastic grid, notated by circles on the image below, which describes the image behavior around a given pixel. It is the result of a convolution of the image with a Gabor filter, which is used to detect shapes and to extract features using image processing. These Gabor jets contain information of multiple orientations and frequencies for each node.

While performing face recognition on a new facial image, the image is matched with training set and best match indicates the identity of the person.

Figure 2.5 The face model constructed by EBGM: (a) an image graph, (b) the facial landmarks of a test image detected by EBGM, (c) the image graph of a test image constructed by EBGM. [25]

A Gabor wavelet transform creates a dynamic link architecture that projects the face onto an elastic grid. [A convolution expresses the amount of overlap from functions, blending the functions together.] Recognition is based on the similarity of the Gabor filter response at each Gabor node. This biologically-based method using Gabor filters is a process executed in the visual cortex of higher mammals. The difficulty with this method is the requirement of accurate landmark localization, which can sometimes be achieved by combining PCA and LDA methods.

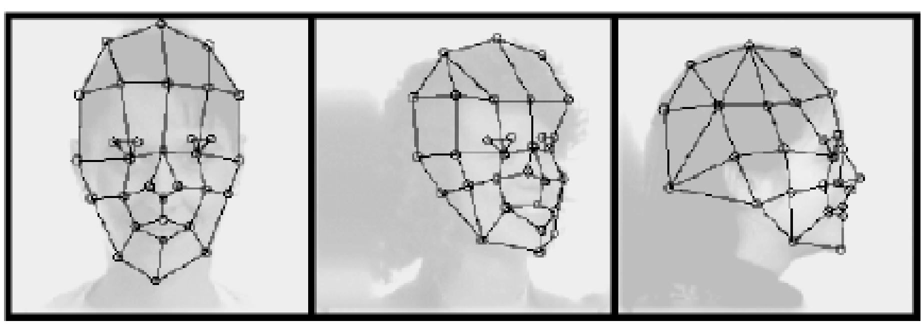


Figure 2.6 Elastic Bunch Map Graphing [25].

2.5.8. Kernel Methods

Kernel methods (KMs) are a class of algorithms for pattern analysis, whose best known element is the support vector machine (SVM). The general task of pattern analysis is to find and study general types of relations (for example clusters, rankings, principal components, correlations, classifications) in general types of data (such as sequences, text documents, sets of points, vectors, images, etc).KMs approach the problem by mapping the data into a high dimensional feature space, where each coordinate corresponds to one feature of the data items, transforming the data into a set of points in a Euclidean space. In that space, a variety of methods can be used to find relations in the data. Since the mapping can be quite general (not necessarily linear, for example), the relations found in this way are accordingly very general. This approach is called the kernel trick. KMs owe their name to the use of kernel functions, that enable them to operate in the feature space without ever computing the coordinates of the data in that space, but rather by simply computing the inner products between the images of all pairs of data in the feature space. This operation is often computationally cheaper than the explicit computation of the coordinates. Kernel functions have been introduced for sequence data, graphs, text, images, as well as vectors.

Algorithms capable of operating with kernels include SVM, Gaussian processes, Fisher's linear discriminant analysis (LDA), principal components analysis (PCA), canonical correlation analysis, ridge regression, spectral clustering, linear adaptive filters and many others.

Because of the particular culture of the research community that has been developing this approach since the mid-1990s, most kernel algorithms are based on convex optimization or eigenproblems, are computationally efficient and statistically well-founded. Typically, their statistical properties are analyzed using statistical learning theory (for example, using Rademacher complexity).

Due to the large appearance changes in human face images, the linear subspace methods may not capture the non-linearity in facial image representation. As a result, the PCA and LDA algorithms have been extended to represent nonlinear mappings in a higher-dimensional space [48].

Computing and storing the new features in this higher-dimensional space becomes very expensive. Thus, the kernel trick is used for computational efficiency as it enables us to obtain the necessary inner products in the higher-dimensional feature space without computing the higher-dimensional feature mapping. Examples of kernel methods are Kernel Eigenfaces and Kernel Fisher faces [49]. Kernel functions can be used without having to form an explicit high-dimensional mapping as long as kernels form an inner product space in this higher dimensional mapping and satisfy Mercer's theorem [50]. A number of papers combining linear subspace methods with the kernel trick including Kernel Direct LDA (KDLDA) [51], Kernel LDA (KDA) [52] or Kernel Fisher's Analysis (KFA), Kernel PCA (KPCA) [53], and Kernel ICA (KICA) [54] have been applied in face recognition showing improved performance over linear approaches.

2.5.9. Tensor faces

Facial images are very much affected by multiple factors such as variations across people, facial expressions, pose changes and lightning conditions. The Tensor faces method [55] is proposed to model the variations of these factors by a multilinear framework. Tensors, which are higher-order extensions of matrices, allow us to construct multilinear models so as to analyze multiple factors of these facial variations. Lathauwer et al. [56] proposed Higher-Order Singular Value Decomposition (HOSVD) for tensor decomposition, which is an extension of Singular Value Decomposition (SVD) for matrix decomposition. Vasilescu et al. [55] introduced the idea of tensor decomposition into the area of computer vision and proposed Tensor faces, a higher-order extension of the Eigenfaces method. To analyze multiple factors of the facial variations, the higher-order extensions of matrices known as Tensors, helps in constructing multilinear models. Tensor consists of training images are analyzed to obtain the basis of each facial factor (expression, pose, etc.) in the training images.

2.5.10. Manifolds

Learning the similarity among data points is one of the key concepts for the analysis of face images. In the previous work of face image analysis using manifold learning methods, it has been shown that face images lie on a manifold [57] [58] [59] [60]. Also, it has been demonstrated that the variation of a certain facial factor such as various poses or expressions makes a sub-manifold in the manifold structure [59]. So, it is helpful to detect and analyze the underlying manifold structure in the distribution of facial image samples. Traditional methods such as PCA and LDA often see only the Euclidean structure, so they fail to discover the underlying structure if the data lies on a nonlinear manifold. The analysis of manifolds reveals the characteristics of the data distribution and can be applied for dimensionality reduction. Thus, to discover the nonlinear structure of manifolds, manifold learning techniques have been proposed [57] [59] [62]. In many real-world classification problems, the local manifold structure is more important than the global Euclidean structure. Thus, manifold learning techniques often use adjacency information among data samples to preserve the local manifold structure. By manifold learning techniques, neighboring points should still be in close proximity after mapping, and the points far from each other should still be far from each other in the new mapping.

2.6 Gabor Filter

For the face recognition feature extraction is a prime and trivial problem. To overcome this in using stipulated facial features that sometimes hinder recognition, bank of Gabor filters are used for automatically extracting all the facial features. The system does not need any preprocessing of input face image and is not based on locating specific and prominent facial features. The system is based on locating all feature points on the face which contain high energy areas. The main advantage of this approach is that if prominent facial features are occluded due to facial hair, glasses or expressions even then it can still recognize the face optimally basing on the features extracted from other regions of the face. The processing of facial images by Gabor filters is chosen due to its biological relevance and technical properties and has similar characteristics to those of the human visual system. Gabor filters are principally band pass filters which are used for feature extraction, texture analysis and texture segmentation, target detection, fractal dimension management, document analysis, edge detection, retina identification, image coding and image representation.

A Gabor filter can be viewed as a sinusoidal plane of particular frequency and orientation, modulated by a Gaussian envelope. It can be written as:

$$h(x, y) = s(x, y) * g(x, y) \quad (2.4)$$

Where, $s(x, y)$ is a complex sinusoid, and

$g(x, y)$ is a 2-D Gaussian shaped function known as envelope, x, y are the pixel coordinates.

The Gabor filter bank is made by tuning the filter with a specific band of spatial frequency and orientation by appropriately selecting the filter parameter. The Gabor representation of a face image is computed by convolving it with the Gabor filter.

The convolved image is used for training the neural network. Gabor filters yields robustness against expression, illumination and small rotations.

2.7 Comparison of different techniques:

S.No.	Technique	Author	Advantages	Disadvantages
1.	Eigenfaces (PCA)	M. Turk and A. Pentland.[33]	(a) Reduction of number of variables, by combining two or more variables into a single factor. (b) It can be used to identify hidden dimensions or constructs which may not be apparent from direct analysis	(a)PCA is good for data representation but not necessarily for class discrimination. (b) It is impossible to pick the proper rotation using PCA alone.

				(c) PCA works well when the lightening variation is small.
2.	Linear Discriminant Analysis (LDA) and Fisherfaces	R. O. Duda, P. E. Hart, and D. G. Stork.[34]	(a) It is more suited for finding projections that best discriminate different classes. (b) The projection vectors from Fisherfaces maximizes the objective function. (c) Powerful tool used for data reduction and feature extraction.	(a) The optimality criteria are not directly related to the classification ability of the obtained feature representation. (b) It is not accurate enough to describe subtleties of original manifolds in the original image space due to their limitations in handling nonlinearity in face recognition.
3.	LDA variants	H. Yu and J. Yang [36], J. Lu, K. N. Plataniotis and A. N. Venetsanopoulos [37], W. Zheng, C. Zou, and L. Zhao. [38].	(a) LDA works gives better accuracy in facial expression. (b) To evaluate the performance of our boosted hybrid method, we compare it to state-	It is very much dependent on the design decisions which involve the method of subspace analysis,

			of-the-art LDA variants. (c)It maximizes the Fisher's ratio.	varying the dimension of the subspace and choosing the similarity measure.
4.	Independent Component Analysis (ICA)	M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski.[39]	ICA seeks a non-orthogonal basis so that the transformed features are statistically independent.	ICA coefficients cannot be recovered from the transformed coefficients.
5.	Local Feature Analysis (LFA)	P. S. Penev and J. J. Atick [40], P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone [41].	(a) It analyzes the image feature and the effectiveness in analyzing spatial location. (b) It increases the stability of recognition performance under changes in viewpoint, illumination, and expression.	The features used in the LFA method are not sensitive to illumination changes.
6.	Neural Networks (NN) and Support Vector Machines (SVM)	S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back [42], S. Gutta, J. R. J. Huang, P. Jonathon, and H. Wechsler [43], D. Zhang, H. Peng, J. Zhou, and S. K. Pal [44], V.N.Vapnik [45], B. Heisele, P. Ho, and T. Poggio [48].	(a) Nonlinear classification for face detection may be performed by this technique. (b) It finds the hyperplane that maximizes the margin of separation in order to minimize the risk of misclassification not only for the training samples.	It involves huge training set with multiple images in different pose and expression for each person. Hence, increasing the computational time.
7.	Elastic Bunch Graph	L. Wiskott, J. M. Fellous,	(a) It reduces the human	(a) In this

	Matching (EBGM)	N. Kr¨uger, and C. vonder Malsburg.[32]	effort as the face is represented in internal form as face graph containing node for each landmark position with the corresponding extracted features. (b) The system does not require large training set for efficient recognition.	system the eyes should be open as the system aligns the images on the basis of eyes locations. (b) The system is not illumination invariant. The change in lightning conditions deteriorates the results of a system.
8.	Kernel Methods	K. I. Kim, K. Jung, and H. J. Kim, G. Baudat and F. Anouar [54], M. H. Yang [50], K. R. Muller, K. Tsuda S. Mika, G. Ratsch, and B. Scholkopf [51], F. R. Bach and M. I. Jordan [55].	It enables us to obtain the necessary inner products in the higher-dimensional feature space without computing the higher-dimensional feature mapping.	It may not perform well for unseen data even having their more flexibility than the linear methods and over fitting thereof.
9.	Tensorfaces	M. A. O. Vasilescu and D. Terzopoulos [56], L. D. Lathauwer, B. D. Moor, and J. Vandewalle [57].	It is proposed to model the variations of these factors like pose changes, lighting conditions and facial expressions by a multilinear framework.	In Tensorfaces it is difficult to analyze the factors of a new test image when all the factors of the test image are unknown or untrained.
10.	Manifolds	X. He and O. Niyogi [59], J. B. Tenenbaum, V. Silva,	The analysis of manifolds reveals the	In manifold learning

		and J. C. Langford [60], M. Belkin and P. Niyogi [61], X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang [62].	characteristics of the data distribution and can be applied for dimensionality reduction.	techniques, neighboring points should still be in close proximity after mapping.

Table 2.2 Comparison of different techniques

Chapter 3

System Implementation

This chapter is dedicated to the explanation of a wavelet, PCA (Eigen Faces) and neural network method that will be used in face recognition, on which the proposed face recognition system is based.

3.1 Wavelets

A wavelet is a wave-like oscillation with amplitude that starts out at zero, increases, and then decreases back to zero. It can typically be visualized as a "brief oscillation" like one might see recorded by a seismograph or heart monitor. Generally, wavelets are purposefully crafted to have specific properties that make them useful for signal processing. Wavelets can be combined, using a "shift, multiply and sum" technique called convolution, with portions of an unknown signal to extract information from the unknown signal.

The frequency bands or subspaces (sub-bands) are scaled versions of a subspace at scale 1. This subspace in turn is in most situations generated by the shifts of one generating function $\psi \in L^2(R)$, *the mother wavelet*. For the example of the scale one frequency band [1,2] this function is:

$$\psi(t) = 2\sin c(2t) - \sin c(t) = \frac{\sin(2\pi t) - \sin(\pi t)}{\pi t}$$

Several families of wavelets that have proven to be especially useful are included in this toolbox. What follows is an introduction to these wavelet families. To explore wavelet families on your own, check out the **Wavelet Display** tool:

1. Type *wavemenu* from the MATLAB command line. The Wavelet Toolbox Main Menu appears.

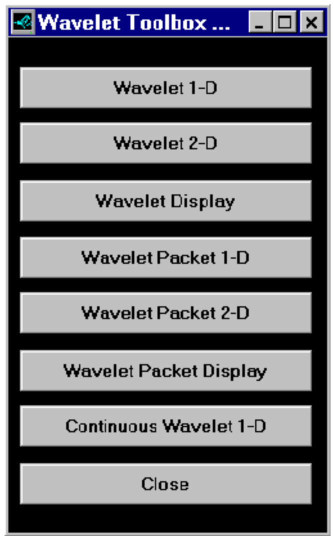


Figure 3.1 Wavelet Toolbox Menu

2. Click on the **Wavelet Display** menu item. The **Wavelet Display** tool appears.
3. Select a family from the **Wavelet** menu at the top right of the tool.
4. Click the **Display** button. Pictures of the wavelets and their associated filters appear.
5. Obtain more information by clicking on the information buttons located at the right.

3.1.1 HAAR

Haar is discontinuous, and resembles a step function. The Haar wavelet is also the simplest possible wavelet. The technical disadvantage of the Haar wavelet is that it is not continuous, and therefore not differentiable. This property can, however, be an advantage for the analysis of signals with sudden transitions, such as extracting features for face recognition and monitoring of tool failure in machines.

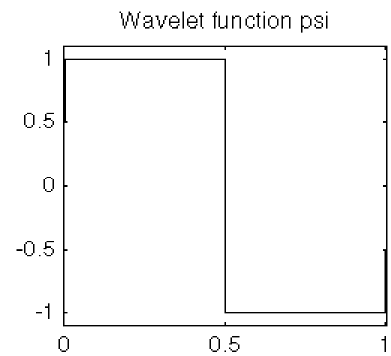


Figure 3.2 Haar Wavelet

3.1.2 DAUBECHIES

Ingrid Daubechies, one of the brightest stars in the world of wavelet research, invented what are called compactly-supported orthonormal wavelets. The names of the Daubechies family wavelets are written dbN, where N is the order, and db the “surname” of the wavelet. The db1 wavelet, as mentioned above, is the same as Haar. Here are the next nine members of the family:

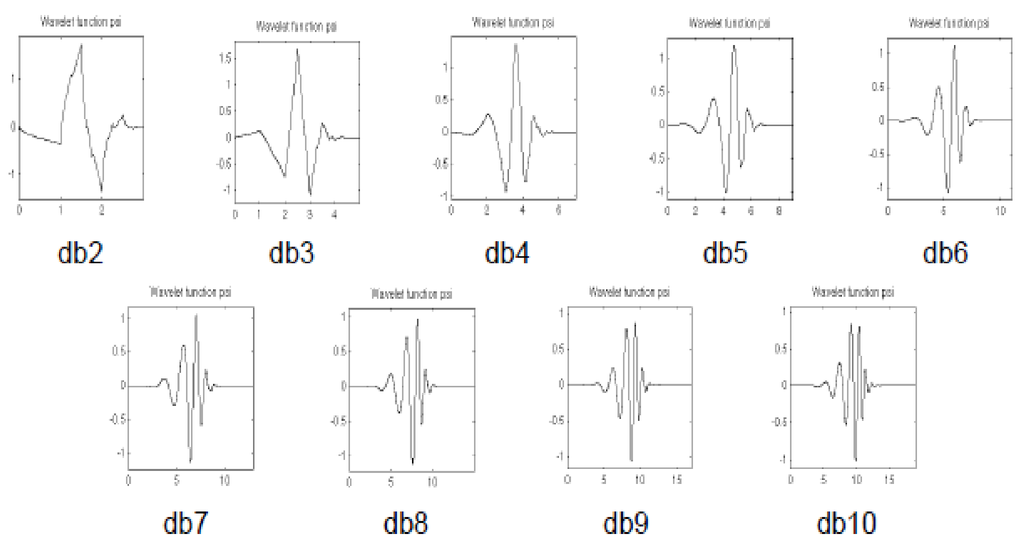
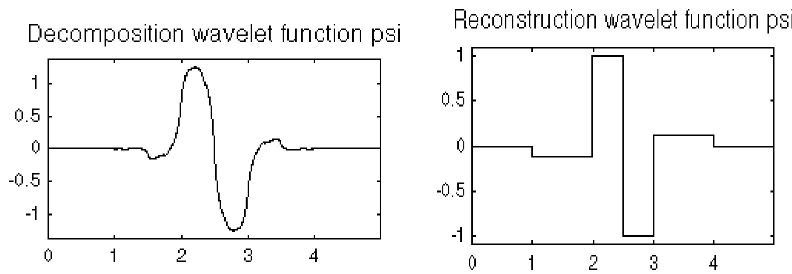


Figure 3.3 Daubechies Wavelet

3.1.3 BIORTHOGONAL

This family of wavelets exhibits the property of linear phase, which is needed for signal and image reconstruction. By using two wavelets, one for decomposition and the other for reconstruction instead of the same single one, interesting properties are derived.

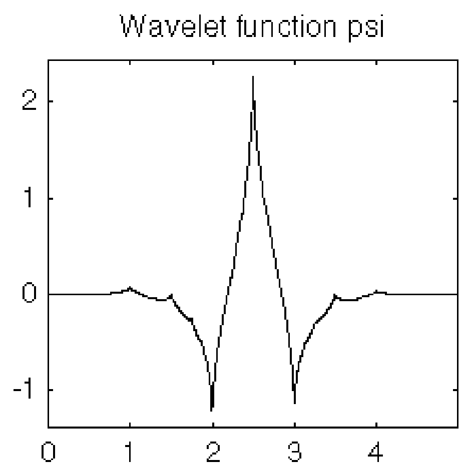


Bior1.3

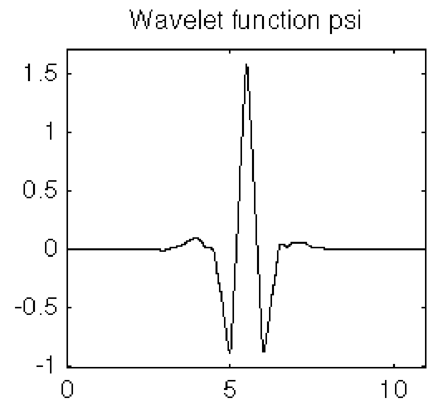
Figure 3.4 Biorthogonal Wavelet

3.1.4 COIFLETS

The wavelet function has 2*N* moments equal to 0 and the scaling function has 2*N*-1 moments equal to 0.The two functions have a support of length 6*N*-1. You can obtain a survey of the main properties of this family by typing *waveinfo ('coif')* from the MATLAB command line.



(a) coif1

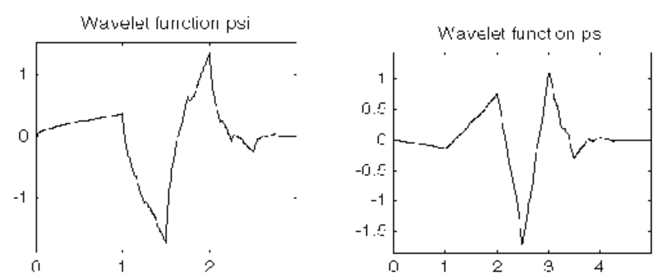


(b) coif2

Figure 3.5 (a) coif1 (b) coif2 Coiflets Wavelet

3.1.5 SYMLETS

The *symlets* are nearly symmetrical wavelets proposed by Daubechies as modifications to the db family. The properties of the two wavelet families are similar.



(a) sym2 (b) sym3

Figure 3.6 (a) sym2 (b) sym3 Symlets Wavelet

3.1.6 MORLET

This wavelet has no scaling function, but is explicit.

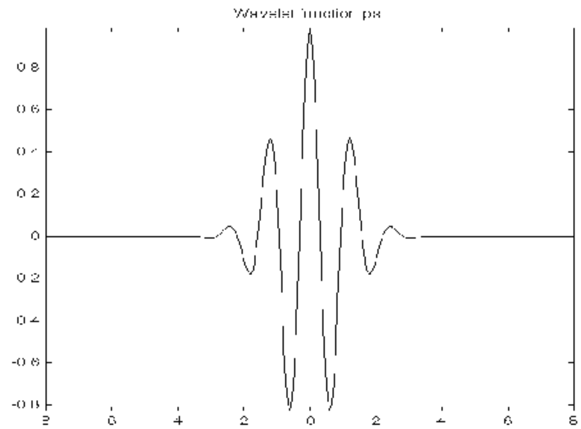


Figure 3.7 Morlet Wavelet

3.1.7 MEXICAN HAT

This wavelet has no scaling function and is derived from a function that is proportional to the second derivative function of the Gaussian probability density function.

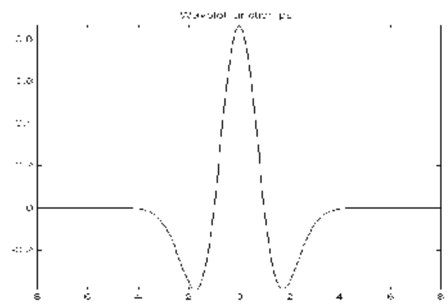


Figure 3.8 Mexican Hat Wavelet

3.1.8 MEYER

The Meyer wavelet and scaling function are defined in the frequency domain.

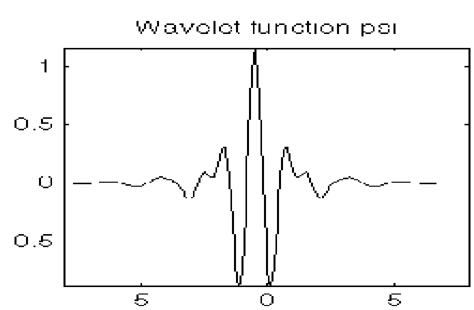


Figure 3.9 Meyer Wavelet

3.1.9 HAAR Wavelet Technique

Haar wavelet basis can be used to represent this image by computing a wavelet transform. Clearly, some information is lost in this averaging process. We need to store some detail coefficients to recover the original four pixel values from the two averaged values.

3.1.9.1 Compression of 2D image with Haar Wavelet Technique

The compression of the 2D image is a 2D generalization of the 1D wavelet. It applies the 1D wavelet transform to each row of pixel values. This operation provides us an average value along with detail coefficients for each row. Next, these transformed rows are treated as if they were themselves an image and apply the 1D transform to each column. The resulting values are all detail coefficients except a single overall average co-efficient. In order to complete the transformation, this process is repeated recursively only on the quadrant containing averages. Figure 3.11 shows a typical 2D DWT, used in image compression, generates the hierarchical structure. Fix a nonnegative threshold value □, and decree that any detail coefficient in the wavelet transformed data whose magnitude is less than or equal to □ will be reset to zero (hopefully, this leads to a relatively sparse matrix), then rebuild an approximation of the original data using this doctored version of the wavelet transformed data.

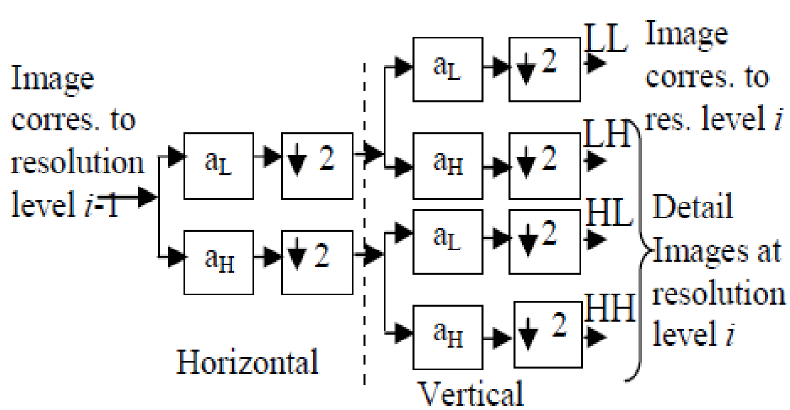


Figure 3.10 One Filter Stage in 2D DWT

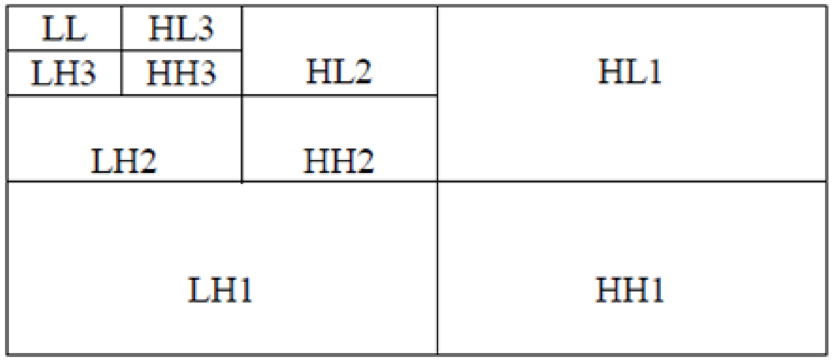


Figure 3.11 Structure of wavelet decomposition

Figure 3.12 shows the Wavelet 2-D Toolbox showing the HAAR level 2 decomposition of the image from the YALE face database. The surprise is that in the case of image data, we can throw out a sizable proportion of the detail coefficients in this way and obtain visually acceptable results. This process is called lossless compression when no information is lost. Figure shows the compressed image at the threshold value of 89.

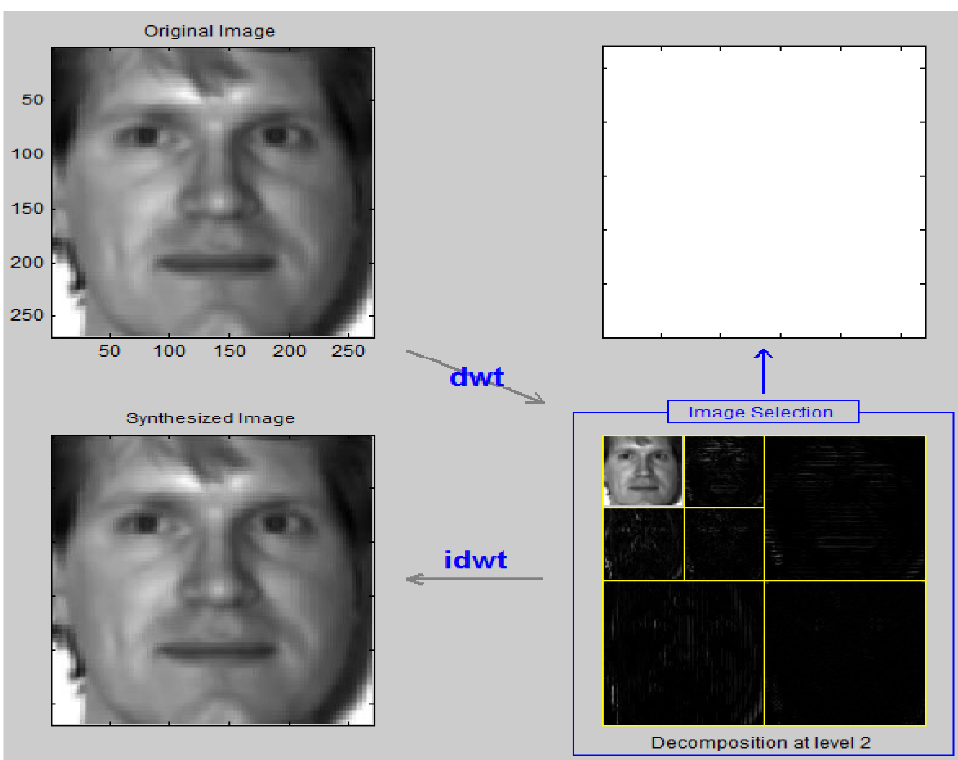


Figure 3.12 HAAR level 2 decomposition of the image from the YALE face database.

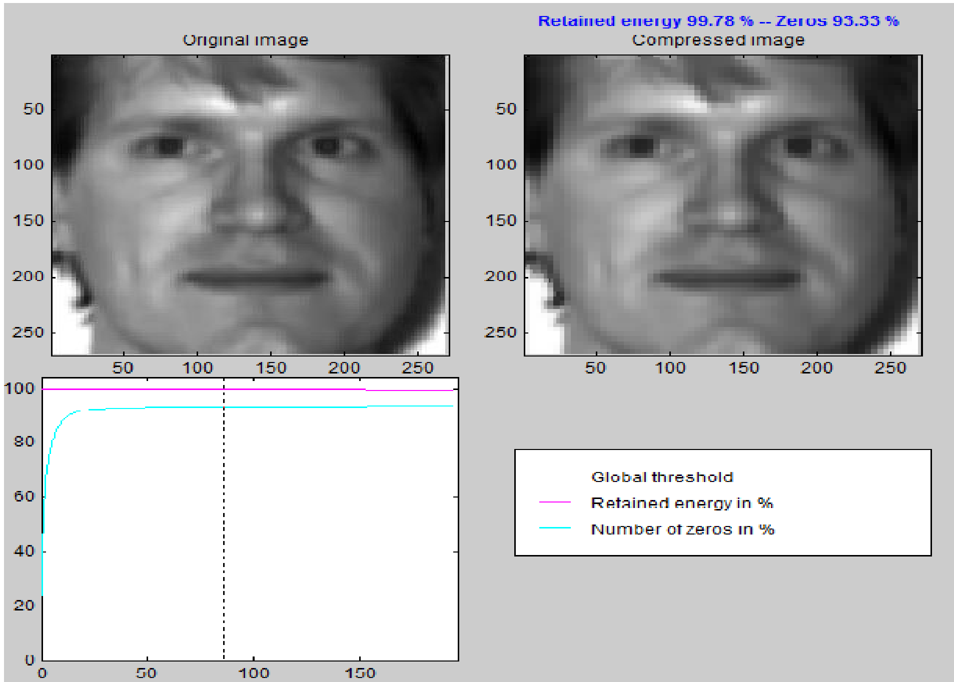


Figure 3.13 Compressed image at the threshold value of 89.

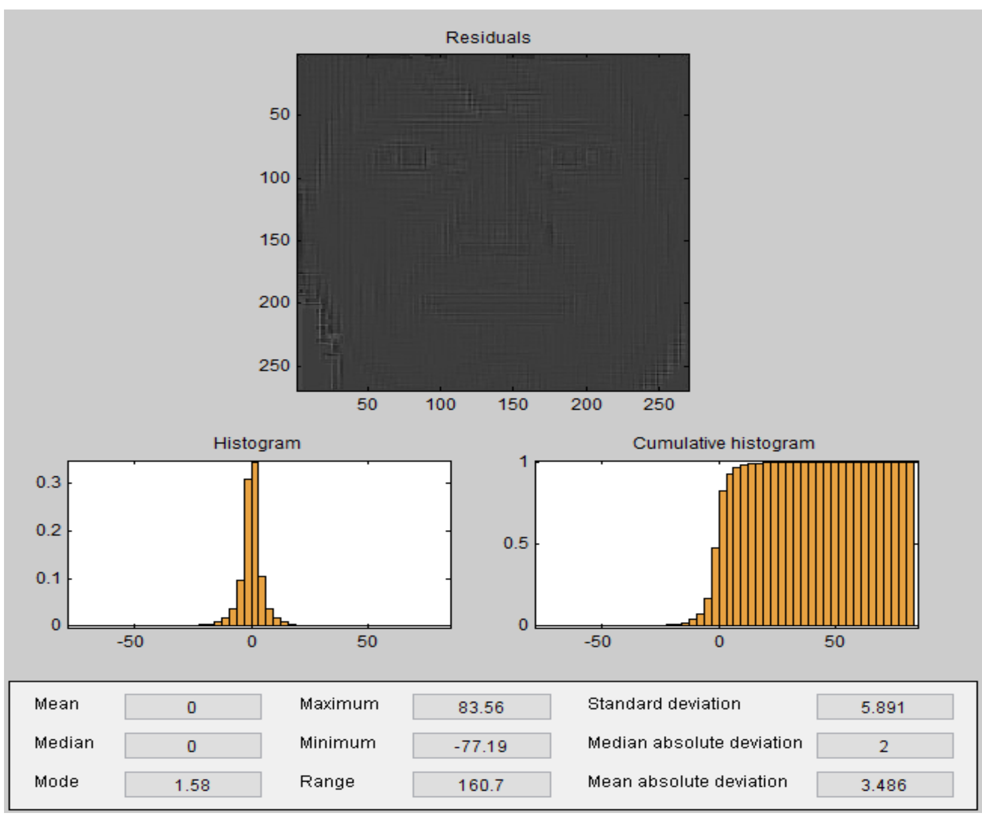


Figure 3.14 Residuals of the image showing all the values.

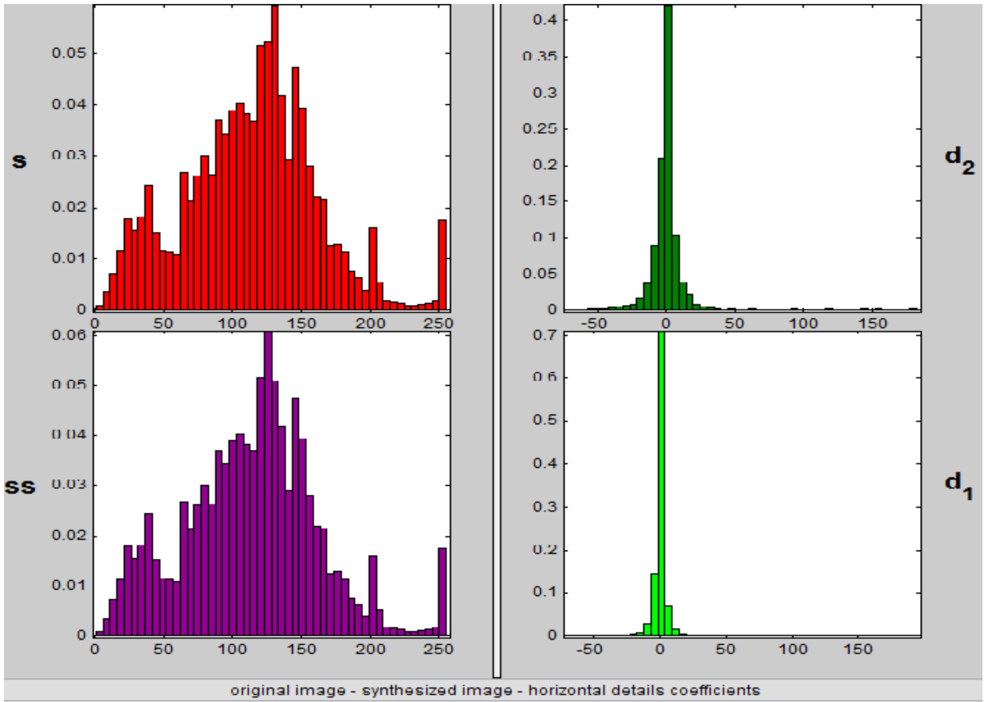


Figure 3.15 Histograms of the original and compressed image

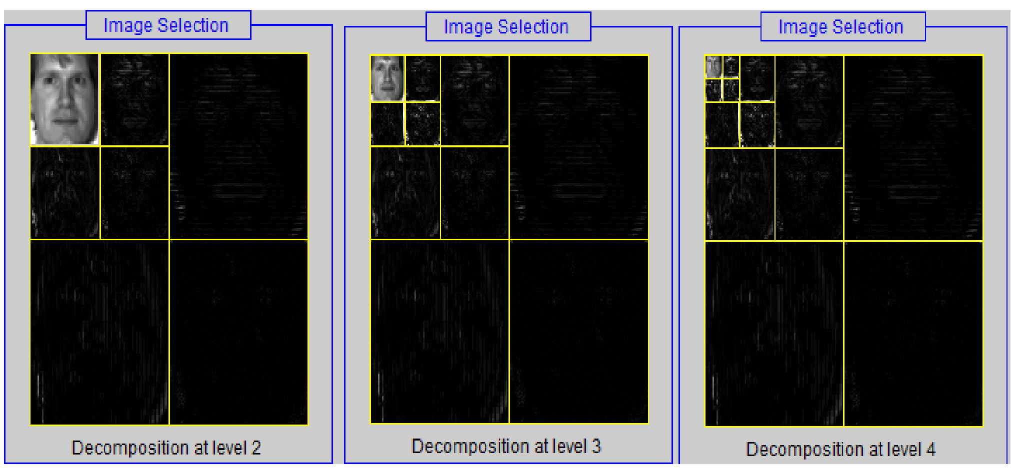


Figure 3.16 Image decomposition at and HAAR 2, 3 and 4 levels.

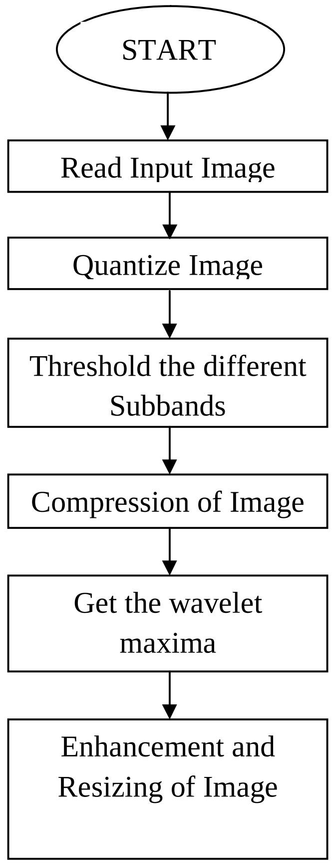


Figure 3.17 Flowchart for the image resizing and enhancement using DWT

The wavelet transform enhances the contrast as well as the edges of face images, in the frequency to facilitate the face detection task. Wavelet transform is better than the histogram equalization as the later only enhances the image pixel gray-level contrast in the spatial domain. With the Wavelet Transform, face detection is much more efficient under the wide range of illumination changes.

3.2 Principle Component Analysis

Much of the previous work on automated face recognition has ignored the issue of just what aspects of the face stimulus are important for face recognition. This suggests the use of an information theory approach of coding and decoding of face images, emphasizing the significant local and global features. Such features may or may not be directly related to our intuitive notion of face features such as the eyes, nose, lips, and hair.

The principal components of the distribution of faces, or the eigenvectors of the covariance matrix of the set of face images, treating an image as point (or vector) in a very high dimensional space is sought. The eigenvectors are ordered, each one accounting for a different amount of the variation among the face images.

These eigenvectors can be thought of as a set of features that together characterize the variation between face images. Each image location contributes more or less to each eigenvector, so that it is possible to display these eigenvectors as a sort of ghostly face image which is called an "*eigenface*".

Sample face images, the average face of them, eigenfaces of the face images and the corresponding eigenvalues are shown in Figure 3.1, Figure 3.2, Figure 3.3, and Figure 3.4 respectively. Each eigenface deviates from uniform gray where some facial feature differs among the set of training faces. Eigenfaces can be viewed as a sort of map of the variations between faces.

Figure 3.18 Sample Faces

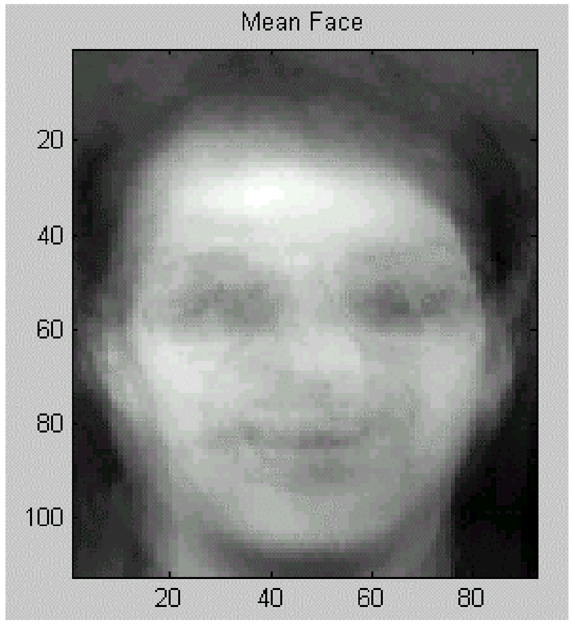


Figure 3.19 Average face of the Sample Faces

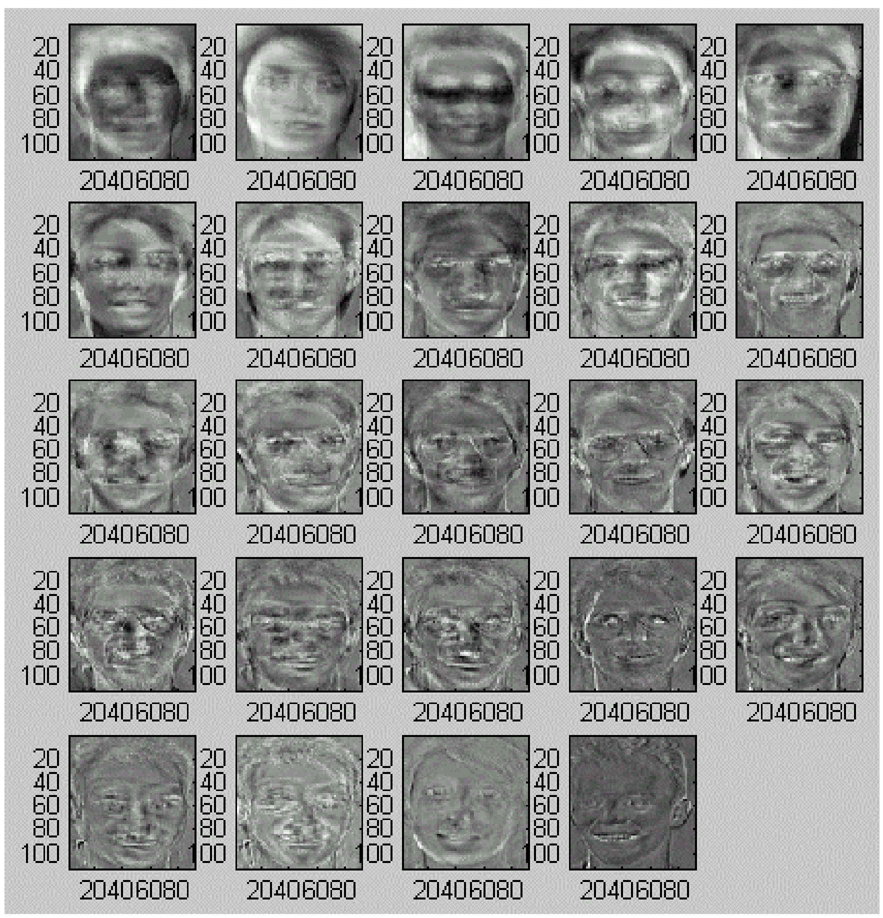


Figure 3.20 Eigen Faces of the Sample Faces

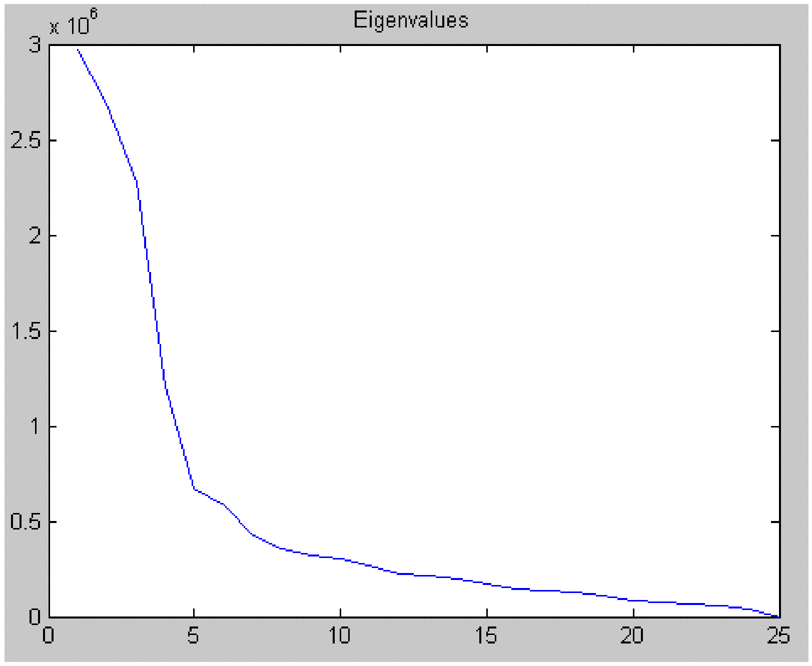


Figure 3.21 Eigenvalues corresponding to eigenfaces

Each individual face can be represented exactly in terms of a linear combination of the eigenfaces. Each face can also be approximated using only the "best" eigenfaces, those that have the largest eigenvalues, and which therefore account for the most variance within the set of face images. As seen from the Figure 3.4, the eigenvalues drops very quickly, that means one can represent the faces with relatively small number of eigenfaces. The best M eigenfaces span an M-dimensional subspace which we call the "face space" of all possible images.

Figure 3.22 Reconstruction of First Image with the number of Eigenfaces.

3.3 Computation of the eigenfaces

Step 1: obtain face images I_1, $I_{2,}$..., I_M (training faces)

(Very important: the face images must be cantered and of the same size)

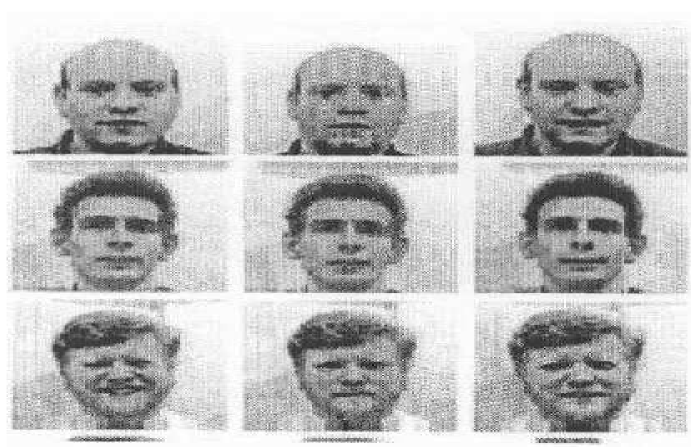


Figure 3.23 Centered face Images

Step 2: Represent every image I_i as a vector Γ_i.

Step 3: Compute the average face vector Ψ:

$$\Psi = \frac{1}{M}\sum_{i=1}^{M}\Gamma_i$$

Step 4: Subtract the mean face:

$$\Phi_i = \Gamma_i - \Psi$$

Step 5: compute the covariance matrix C:

$$C = \frac{1}{M}\sum_{n=1}^{M}\Phi_n\Phi_n^T = AA^T \quad (N^2 X N^2 \text{ Matrix})$$

$$Where\ A = [\Phi_1 \Phi_2 \ldots\ldots \Phi_M] \quad (N^2 X\ M \text{ Matrix})$$

Step 6: Compute the eigenvectors u_i of AA^T.

Step 6.1: Consider the matrix AA^T. (M x M matrix)

Step 6.2: compute the eigenvectors v_i of AA^T.

$$AA^T v_i = \mu_i v_i$$

Step 6.3: compute the M best eigenvectors of AA^T: $ui = Avi$

(**Important:** Normalize ui such that $||u_i|| = 1$)

Step 7: Keep only K eigenvectors (corresponding to the K largest eigenvalues).

3.3.1 Representing faces onto this basis

- Each face (minus the mean) Φ_i in the training set can be represented as a linear combination of the best K eigenvectors:

$$\Phi_i - mean = \sum_{i=j}^{K} w_j u_j, \quad (w_j = u_j^T \Phi_i)$$

(we call the u_j 's $eigenfaces$)

Figure 3.24 Eigen Faces

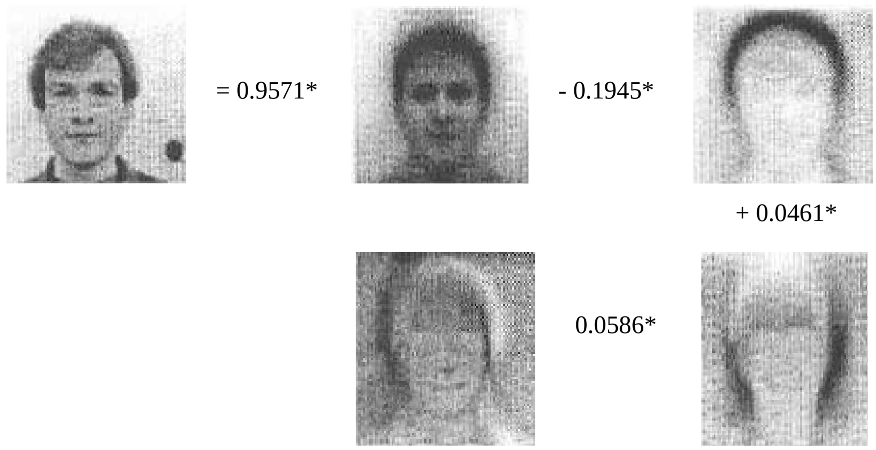


Figure 3.25 Calculation of Eigen Faces

- Each normalized training face Φ_i is represented in this basis by a vector:

$$\Phi_i = \begin{bmatrix} w_1^i \\ w_2^i \\ w_k^i \end{bmatrix} \quad i=1, 2,\ldots,M$$

3.3.2 Face Detection Using Eigenfaces

- Given an unknown image Γ

Step 1: Compute $\Phi = \Gamma - \Psi$

Step 2: Compute $\hat{\Phi} = \sum_{i=1}^{K} w_i u_i \ (w_i = u_i^T \Phi)$

Step 3: compute $e_d = \|\Phi - \hat{\Phi}\|$

Step 4: if $e_d < T_d$, then Γ is a face.

- The distance e_d is called distance from face space (dffs)

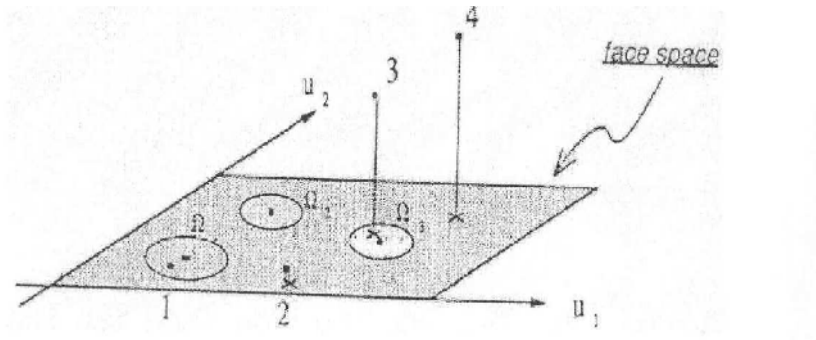


Figure 3.26 Calculation of free space

- **Time requirements**
  - About 400 msec (Lisp, Sun4, 128x128 images)
- **Applications**
  - Face detection, tracking, and recognition
- **Problems**
  - Background (deemphasize the outside of the face, e.g., by multiplying the input image by a 2D Gaussian window centered on the face)
  - Lighting conditions (performance degrades with light changes)
  - Scale (performance decreases quickly with changes to the head size)
    * Multiscale eigenspace
    * Scale input image to multiple sizes)
  - Orientation (performance decreases but not as fast as with scale changes)
    * Plane rotations can be handled
    * Out-of-plane rotations more difficult to handle
- **Experiments**
  - 13 subjects
  - 6 Different mood conditions
  - Total number of images: 78

Figure 3.27 Yale face Database

3.3 Neural Networks

3.3.1. Introduction

Neural networks are composed of simple elements operating in parallel. The neuron model shown in Figure 3.10 is the one that widely used in artificial neural networks with some minor modifications on it.

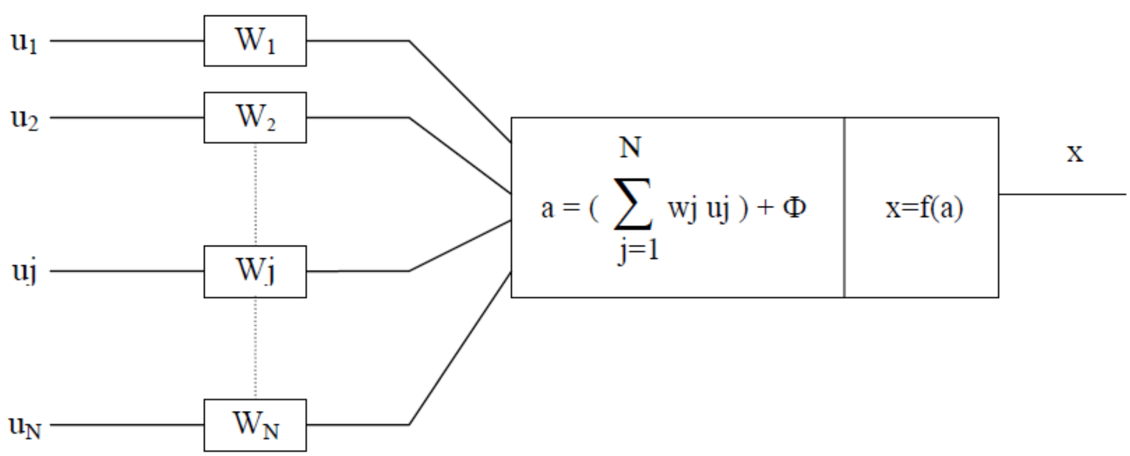


Figure 3.28 Artificial Neuron

The artificial neuron given in this figure has N input, denoted as $u_1, u_2, \dots u_N$. Each line connecting these inputs to the neuron is assigned a weight, which is denoted as $w_1, w_2, \dots, w_N$ respectively. The threshold in artificial neuron is usually represented by Φ and the activation is given by the formula:

$$a = (\sum_{j=1}^{n} w_j u_j) + \Phi \quad (3.1)$$

The inputs and weight are real values. A negative value for a weight indicates an inhibitory connection while a positive value indicating excitatory one. If Φ is positive, it is usually referred as bias. For its mathematical convenience (+) sign is used in the activation formula. Sometimes, the threshold is combined for simplicity into the summation part by assuming an imaginary input u_0 = +1 and a connection weight $w0$ = Φ. Hence the activation formula becomes

$$a = (\sum_{j=1}^{n} w_j u_j) + \Phi \quad (3.2)$$

The vector notation

$$A = w^T u + \Phi \quad (3.3)$$

is useful for expressing the activation for a neuron.

The neuron output function *f (a)* can be:

Identity Function

The function is given by:

$$f(x) = x; \quad \forall x \quad (3.4)$$

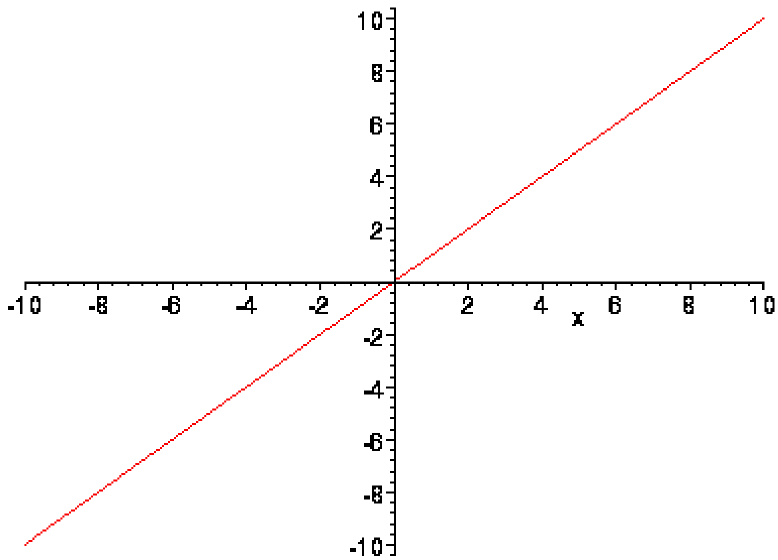


Figure 3.29 Identity Function

Step Function

The step function is given by.

$$f(x) = 0 \quad x \leq 0 \tag{3.5}$$

$$f(x) = 1 \quad x > 0 \tag{3.6}$$

This is also called the heaviside function. Another common variation is for it to take on values -1 and +1 as shown figure 3.30.

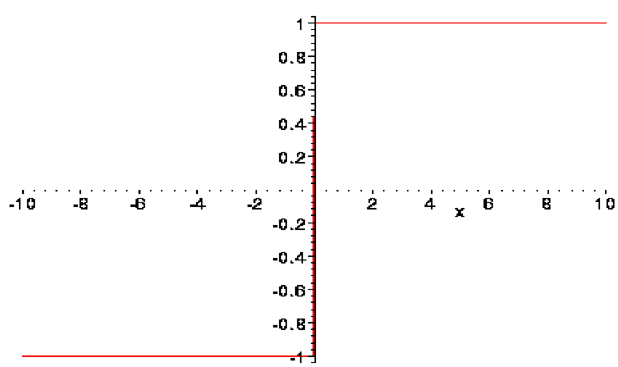


Figure 3.30 Step Function

Sigmoidal Function

The sigmoidal functions are usually S- shaped curves. The hyperbolic and logistic functions are commonly used. These are used in multilayer nets like back propagation network, radial basis function network etc..There are two main types of sigmoidal functions.

1) Logistic Function (Sigmoid)

The logistic function has the form

$$f(x) = \frac{1}{1+e^{(-\alpha x)}} \quad (3.7)$$

the parameter "a" in the logistic function determines how steep it is.

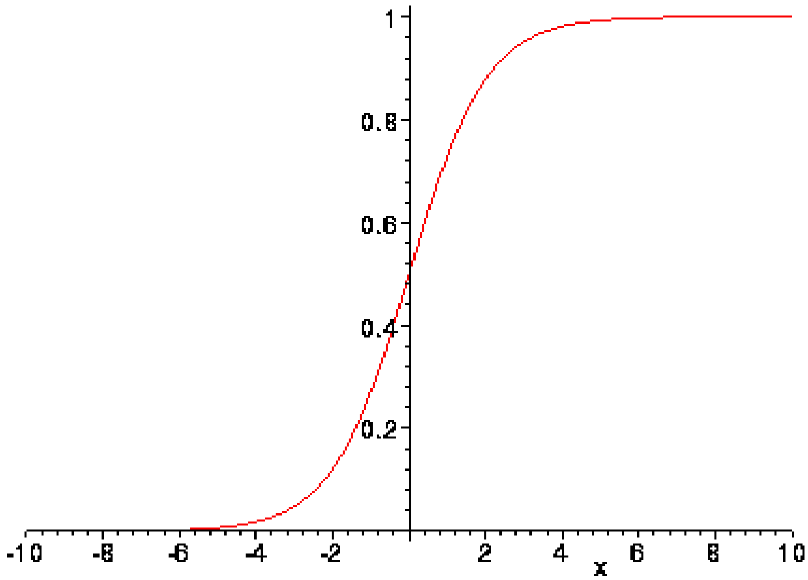


Figure 3.31 Logistic Function

2) Symmetric Sigmoid

The symmetric sigmoid is simply the sigmoid that is stretched so that the y range is 2 and then shifted down by 1 so that it ranges between -1 and 1. If g(x) is the standard sigmoid

$$f(x) = 2g(x) - 1 \quad (3.8)$$

then the symmetric sigmoid is shown in figure 3.14. The symmetric sigmoid differs from the hyperbolic tangent by a constant factor.

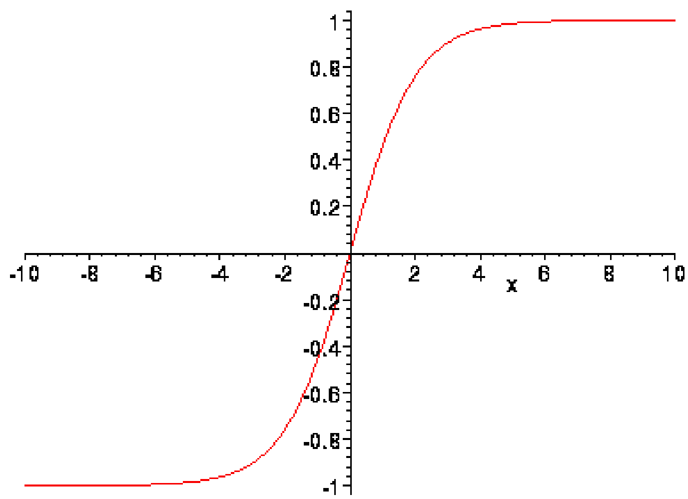


Figure 3.32 Symmetric Sigmoid

Radial Basis Functions

A radial basis function is simply a Gaussian

$$f(x) = e^{(-\alpha x^2)} \qquad (3.9)$$

It is called local because, unlike the previous functions, it is essentially zero everywhere except in a small region. The figure 3.33 shows the radial biasis function.

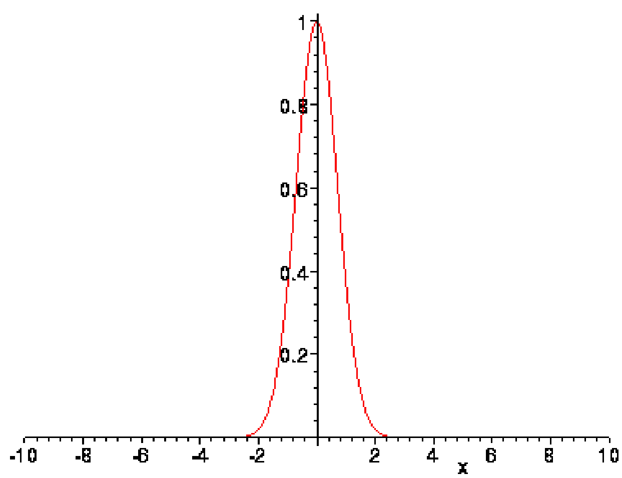


Figure 3.33 Radial Basis Function.

3.3.2 Bias

A bias acts exactly as a weight on a connection from a unit whose activation is always 1. Increasing the bias increases the net input to the unit.

The bias improves the performance of the neural network. If bias is present, then net input is calculated as,

$$Net = b + \sum_i xw_i \tag{3.10}$$

Where, Net = net input,

b = bias,

x_i = input from neurons *i*,

w_i = Weight of the neuron *i* to the output neuron.

3.3.3 Threshold

The threshold ‘θ’ is a factor which is used in calculating the activation of the given net. Function implementations can be done by adjusting the weights and the threshold of the neuron. Furthermore, by connecting the outputs of some neurons as inputs to the others, neural network will be established, and any function can be implemented by these networks. The last layer of neurons is called the output layer and the layers between the input and output layer are called the hidden layers. There are two types of network architecture: *recurrent and feed forward neural network.*

3.3.4 Recurrent Neural Networks:

The structures, in which connections to the neurons of the same layer or to the previous layers are allowed, are called recurrent networks and shown in Figure 3.34.

A recurrent neural network (RNN) is a class of neural network where connections between units form a directed cycle. This creates an internal state of the network which allows it to exhibit dynamic temporal behaviour. Unlike feedforward neural networks, RNN can use their internal memory to process arbitrary sequences of inputs. This makes them applicable to tasks such as unsegmented connected handwriting recognition, where they have achieved the best known results. A network of neuron-like units, each with a directed connection to every other unit. Each unit has a time-varying real-valued activation. Each connection has a modifiable real-valued weight. Some of the nodes are called input nodes, some output nodes, the rest hidden nodes.

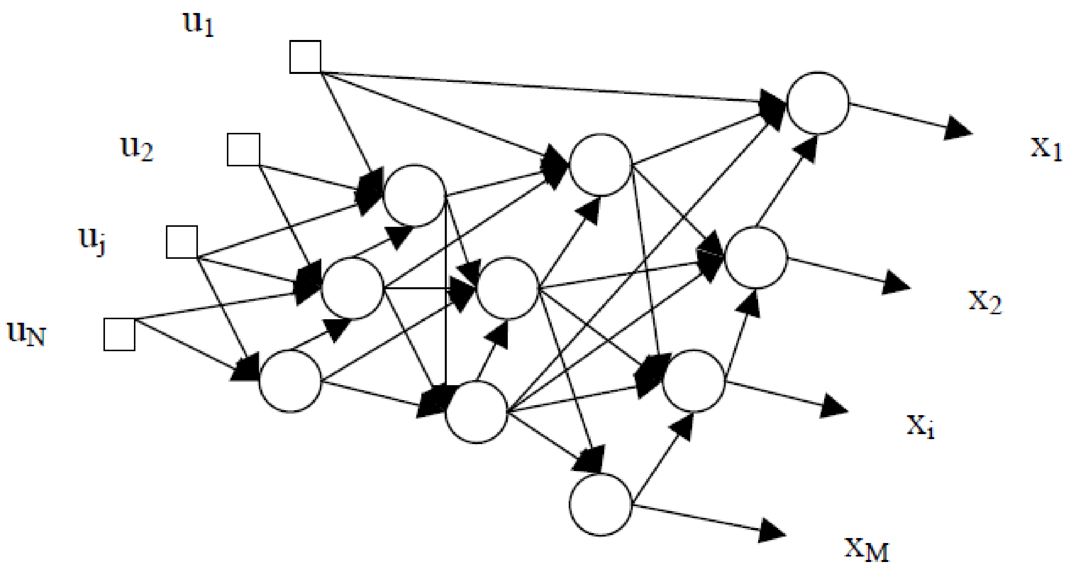


Figure 3.34 Recurrent Neural Networks

3.3.5 Feedforward Neural Networks:

A feedforward neural network is an artificial neural network where connections between the units do not form a directed cycle. In this kind of networks, the neurons are organized in the form of layers. The neurons in a layer get input from the previous layer and feed their output to the next layer. In this kind of networks connections to the neurons in the same or previous layers are not permitted. Figure 3.35 shows typical feedforward neural network.

The feedforward neural network was the first and arguably simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network.

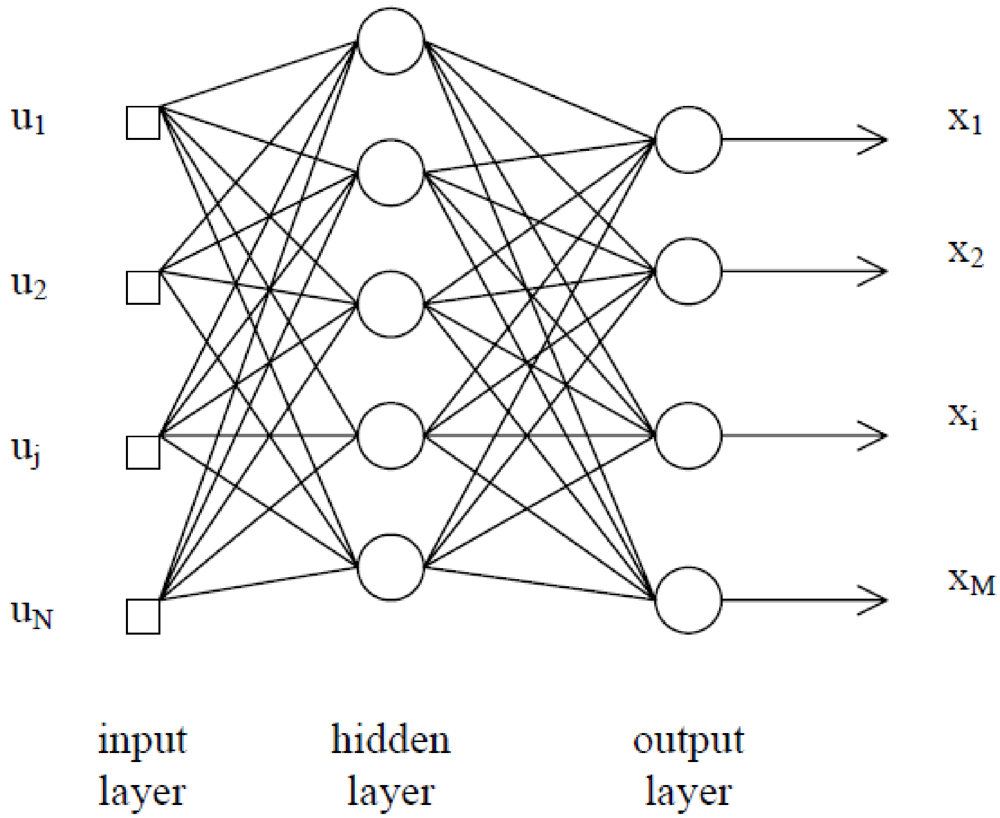


Figure 3.35 Feedforward Neural Networks

For a feedforward network always exists an assignment of indices to neurons resulting in a triangular weight matrix. Furthermore if the diagonal entries are zero this indicates that there is no self-feedback on the neurons. However in recurrent networks, due to feedback, it is not possible to obtain triangular weight matrix with any assignment of the indices.

3.3.6 Back Propagation Algorithm

The training involves four stages:

1. Initialization of weights.
2. Feed forward.
3. Back Propagation of errors.
4. Updating of weights and biases.

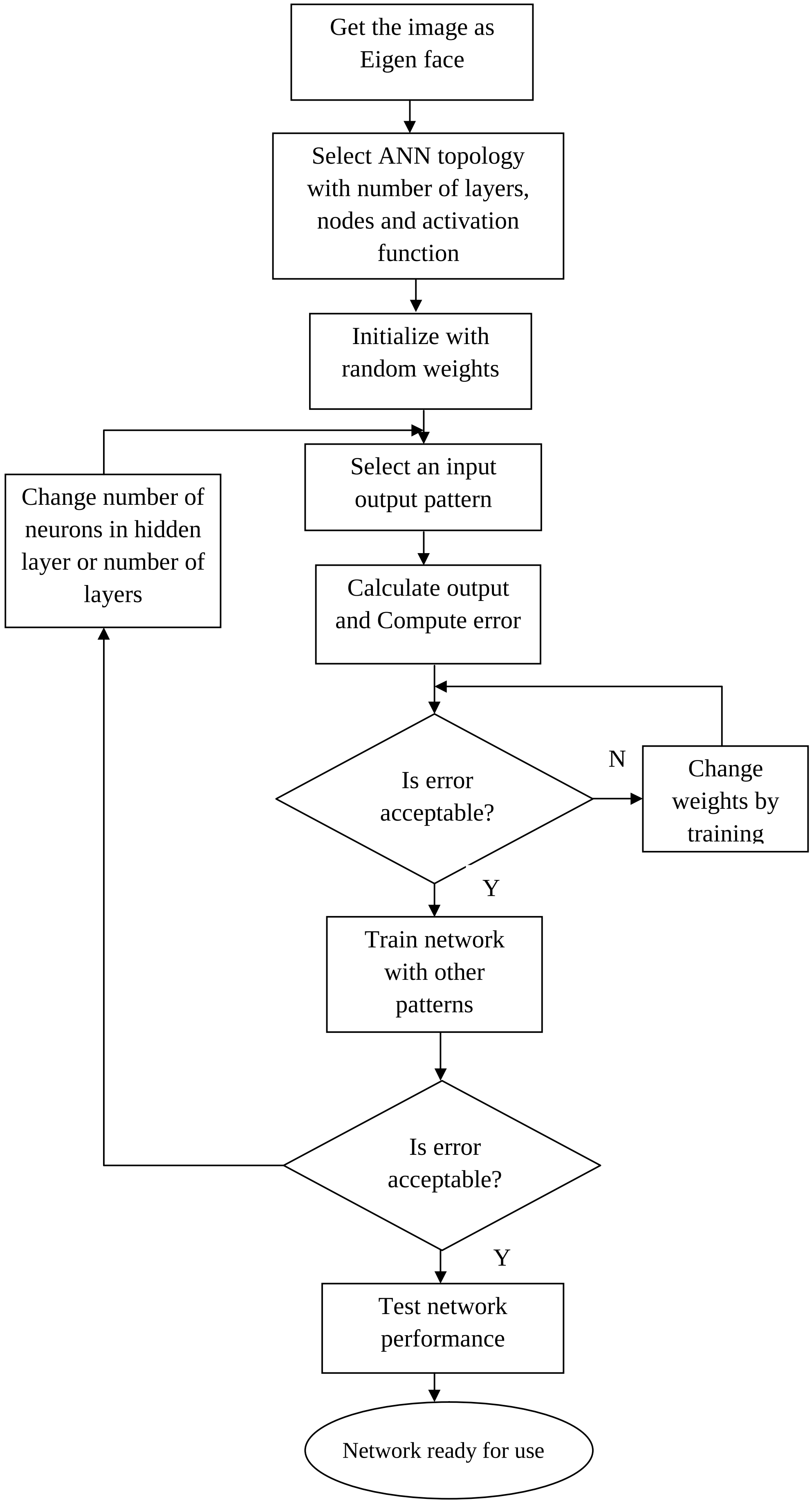


Figure 3.36 Flowchart for Back Propagation Algorithm

Chapter 4

Results and Discussion

4.1 Introduction

Face recognition is an interdisciplinary research area, involving researchers from pattern recognition, computer vision, and graphics, image processing/ understanding, statistical computing and machine learning.

A typical face recognition system is shown in the figure 4.1.1

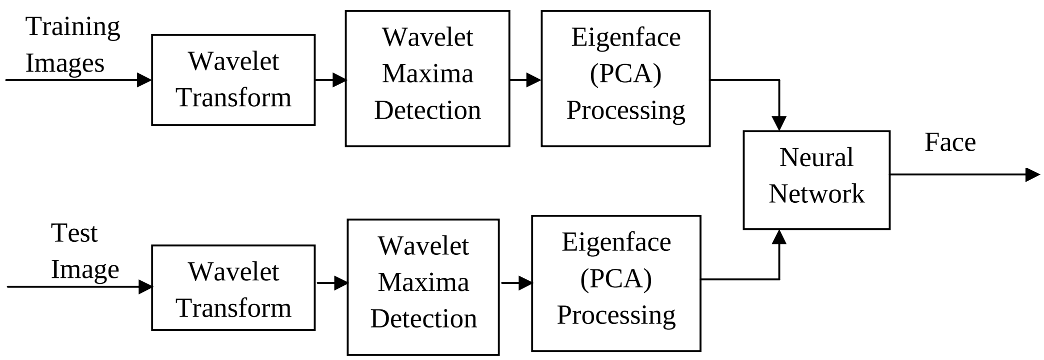


Figure 4.1 Face Recognition Systems

4.1.1 Acquisition of Training Data

The function of this module is to capture the required face image and make it available for the future stage. An acquisition module can get a face image from several different environments. The face image can be an image file that is located on a magnetic disk, it can be captured by a frame grabber or it can be scanned from paper with the help of a scanner. Face images are available from various databases such as Olivetti and Oracle Research Laboratory (ORL), Yale and FERET face databases. Each database has more than one face images with different conditions (expression, illumination etc.), of each individual. The Yale database has been used in the present work. The Yale Face Database contains 78 grayscale images in bmp format of 13 individuals. There are 6 images per subject, one per different facial expression or configuration, center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, and sleepy, surprised, and wink. The whole database is shown in figure 4.2 below.

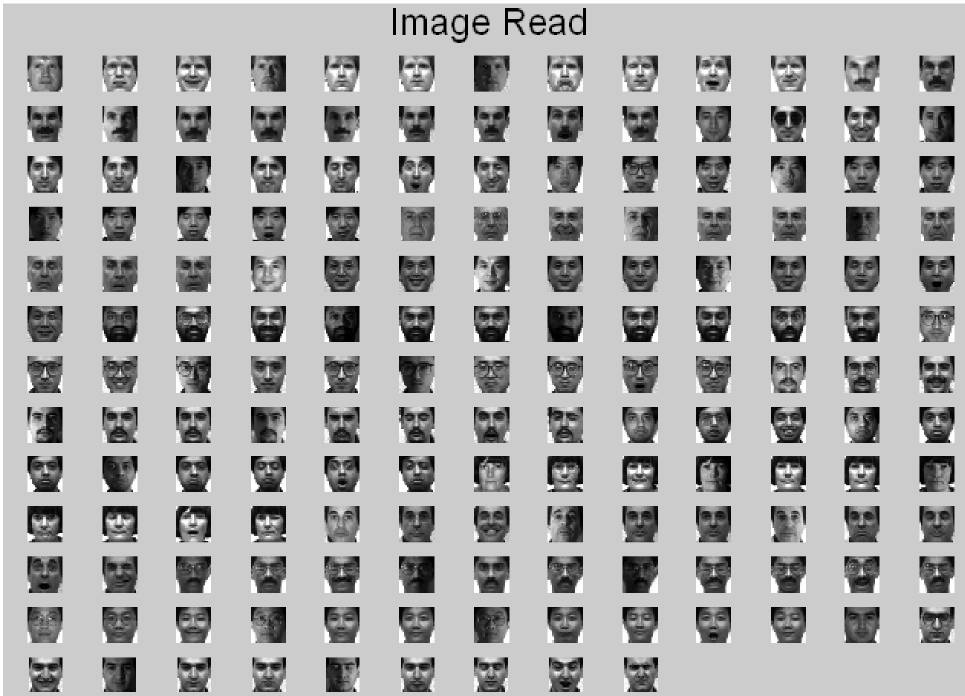


FIGURE 4.2 Yale Face Database

4.1.2 Pre-Processing Module

In this module face images are normalized and if desired, they are enhanced to improve the recognition performance of the system. The preprocessing is done in MATLAB Version 7.5.0.267 (R2007b) using commands of image processing. The pre-processing steps implemented in the present work are explained below:

4.1.2.1 Wavelet Transform

The mathematical background and the advantages of WT in signal processing have been discussed in many research articles. In the proposed system, WT is chosen to be used in image decomposition because:

1. By decomposing an image using WT, the resolution of the sub images are reduced. In turn, the computational complexity will be reduced dramatically by operating on a lower a resolution image. Comparing with the original image

resolution of 128x128, size of the sub-image is reduced by 64 times, and the implies a 64 times reduction in recognition computational load.

2. Under WT, images are decomposed into subbands, corresponding to different frequency ranges. These subbands meet readily with the input requirement for the next major step, and thus minimize the computational overhead in the proposed system.

3. Wavelet decomposition provides the local information in both space domain and frequency domain, while the Fourier decomposition only supports global information in frequency domain.

An image is decomposed into four sub bands as show in figure 4.3. The band LL is a coarser approximation to the original image, bands LH and HL record respectively the changes of the image along horizontal and vertical directions while the HH band shows the higher frequency component of the image. This is the first level decomposition. The decomposition can be further carried out for the LL sub band.

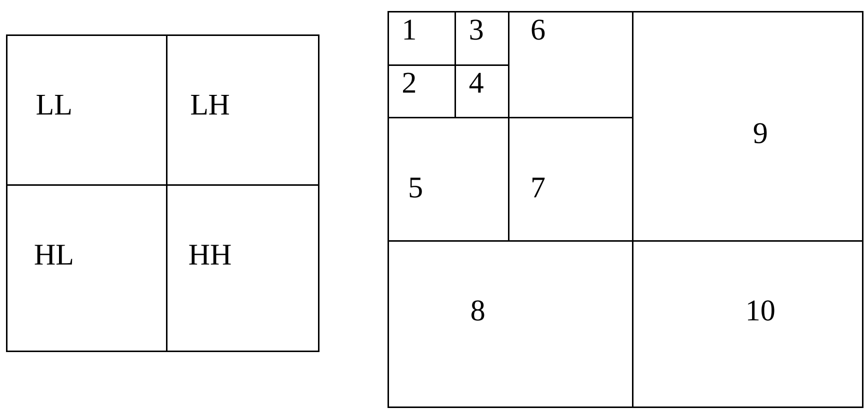


FIGURE 4.3 Wavelet decomposition

Wavelet decomposition provides local information in both space domain and frequency domain. Despite the equal sub band sizes, different sub bands carry different amounts of information. The letter ‘L’ stands for low frequency and the letter ‘H’ stands for high frequency. The left upper band is called LL band because it contains low frequency information in both the row and column directions. The LL band is a coarser approximation to the original image containing the overall information about the whole image. The LH sub band is the result of applying the filter bank column wise and extracts the facial features very well. The HL sub band, which is the result of applying the filter bank row wise, extracts the outline of the face boundary very well. While the HH band shows the high frequency component of the image in non-horizontal, non-vertical directions it proved to be a redundant sub band and was not considered having significant information about the face.

Special hardware is required to make the algorithm work in real time. Thus choosing a wavelet for face detection depends on a lot of trial and error. Discrete Wavelet Transform is recursively applied to all the images in the training data set until the lowest frequency sub band is of size 32x32 pixels i.e. the LH sub band at a particular level or depth of DWT is of size 32x32. The original image's gray scale image is shown in figure 4.4. Here we have used HAAR wavelet instead of Gabor wavelet while calculating wavelet transform.

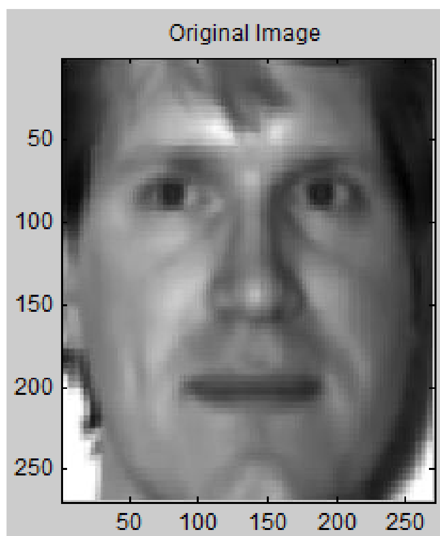


FIGURE 4.4 Original Face image of Yale database

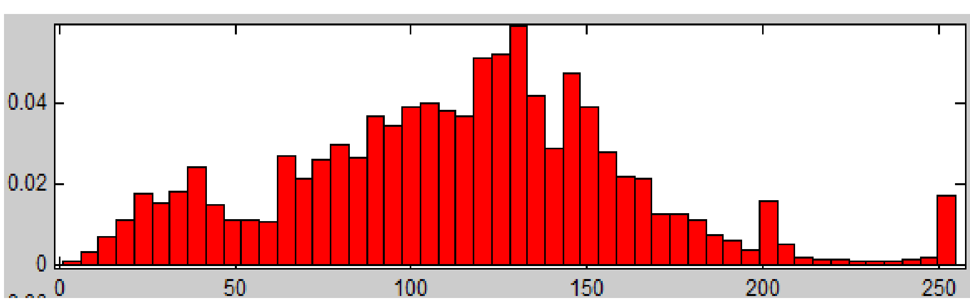


FIGURE 4.5 Histogram of Original Face image of Yale database before DWT

FIGURE 4.6 Decomposed Face image at level 2 after Wavelet Transform

FIGURE 4.7 Synthesized Face image after Wavelet Transform

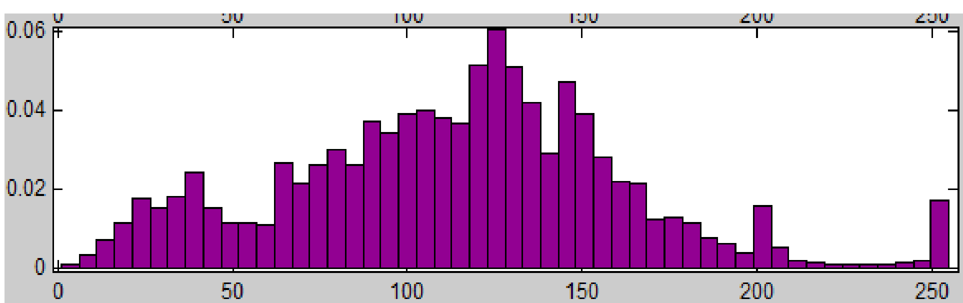


FIGURE 4.8 Histogram of Synthesized image of Yale database after DWT

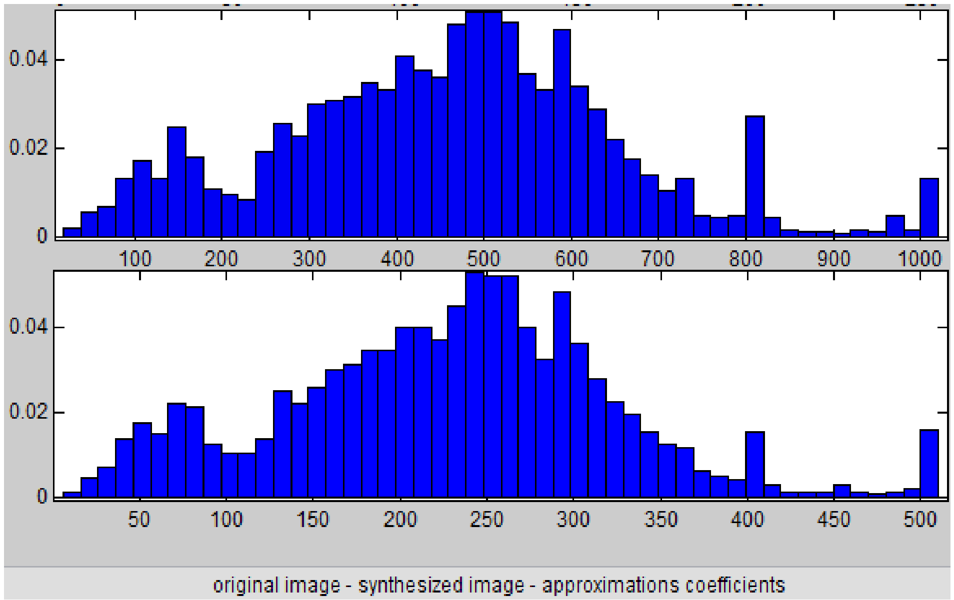


FIGURE 4.9 Approximations Coefficients of original and synthesized image by 2-level HAAR wavelet.

We take the modulus of the wavelet coefficients in the LH sub band. Experiments were performed to go to a resolution even coarser than 32x32. However, it was observed that in certain cases the features would be too close to each other and it was difficult even manually too to separate them.

FIGURE 4.10 The LH sub band of the Face Image.

FIGURE 4.11 The decomposition of the Face Image at the HAAR level 2 in TREE mode.

4.1.2.2 Detection of Wavelet Maxima and De-noising of Image

Our approach to face detection is based on the observation that, in intensity images differ from the rest of the face because of their low intensity. Even if the eyes are closed, the darkness of the eye sockets is sufficient to extract the eye regions. These intensity peaks are well captured by the wavelet coefficients. Thus, wavelet coefficients have a high value at the coordinates surrounding the eyes. We then detect the wavelet maxima or the wavelet peaks in this LH sub band of resolution 32x32.

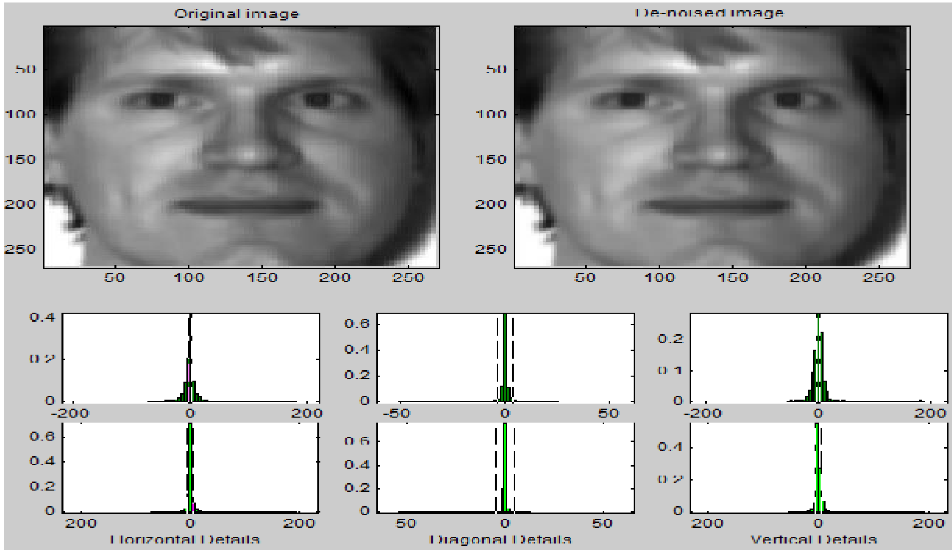
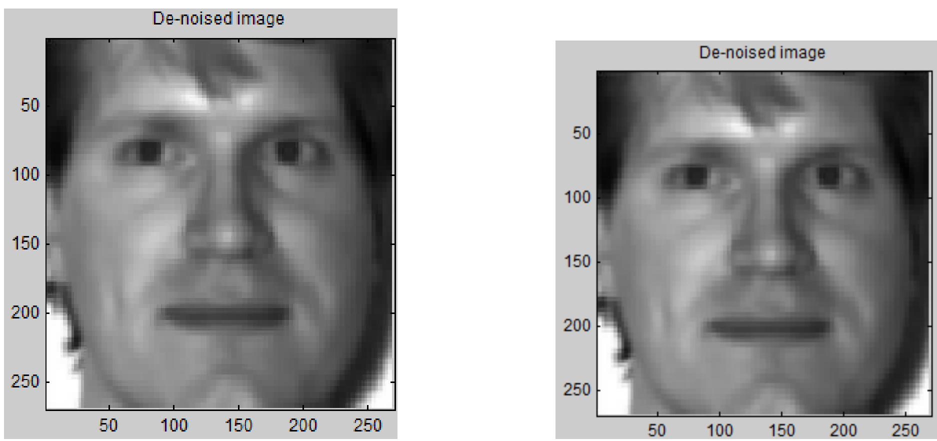


Figure 4.12 De-noised images at the Threshold of 2.1 and the Horizontal, Diagonal and Vertical Coefficients.



Threshold Level =10

Threshold level= 30

Figure 4.13 De-noised image at the Threshold level of 10 and 30.

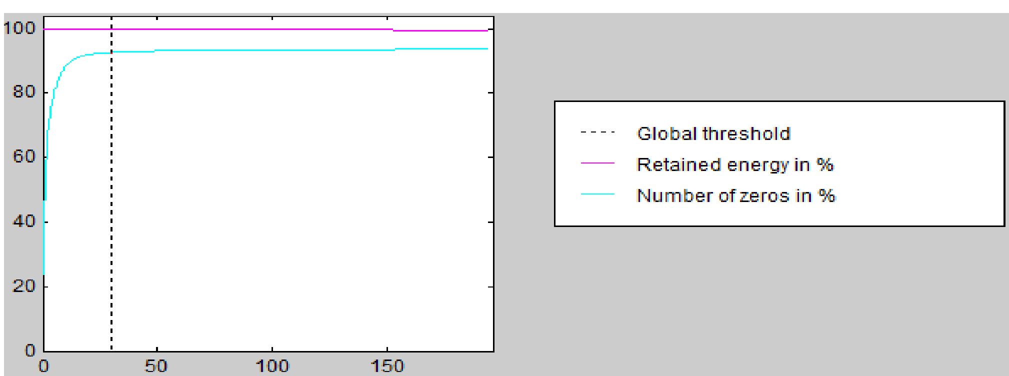


Figure 4.14 Graph showing the relation between global threshold, retained energy and number of zeros.

Global Threshold	Retained Energy	Number of zeroes
5	**99.99%**	**80.26%**
10	**99.96%**	**88.43%**
20	**99.92%**	**99.91%**
30	**99.89%**	**92.60%**

Table 4.1 Retained energy and number of zeroes

4.1.3 Eigenfaces (PCA) processing

The principal components of the distribution of faces, or the eigenvectors of the covariance matrix of the set of face images, treating an image as point (or vector) in a very high dimensional space is sought. The eigenvectors are ordered, each one accounting for a different amount of the variation among the face images.

These eigenvectors can be thought of as a set of features that together characterize the variation between face images. Each image location contributes more or less to each eigenvector, so that it is possible to display these eigenvectors as a sort of ghostly face image which is called an "eigenface".

4.1.3.1 Image Size Normalization

The size of the image is reduced to 25×25 from 100×100 sized image using nearest neighbor interpolation. If the resizing is not done on the images than extra nodes are to be created in the neural network which will greatly reduce the efficiency of the neural network during training.

Codes for the Image resize in MATLAB:

```
function newimg = im_resize(img,nw,nh)
%IM_RESIZE Resize an image using bicubic interpolation
%
%	NEWIMG = IM_RESIZE(IMG,NW,NH) Given input image IMG,
%	returns a new image NEWIMG of size NWxNH.
```

```
% Matthew Dailey 2000

 if nargin ~= 3
   error('usage: im_resize(image,new_wid,new_ht)');
 end;
 ht_scale = size(img,1) / nh;
 wid_scale = size(img,2) / nw;
 newimg = interp2(img,(1:nw)*wid_scale,(1:nh)'*ht_scale,'cubic');
```

(a) (b)

Figure 4.15 (a) Original Image (b) Resized Image

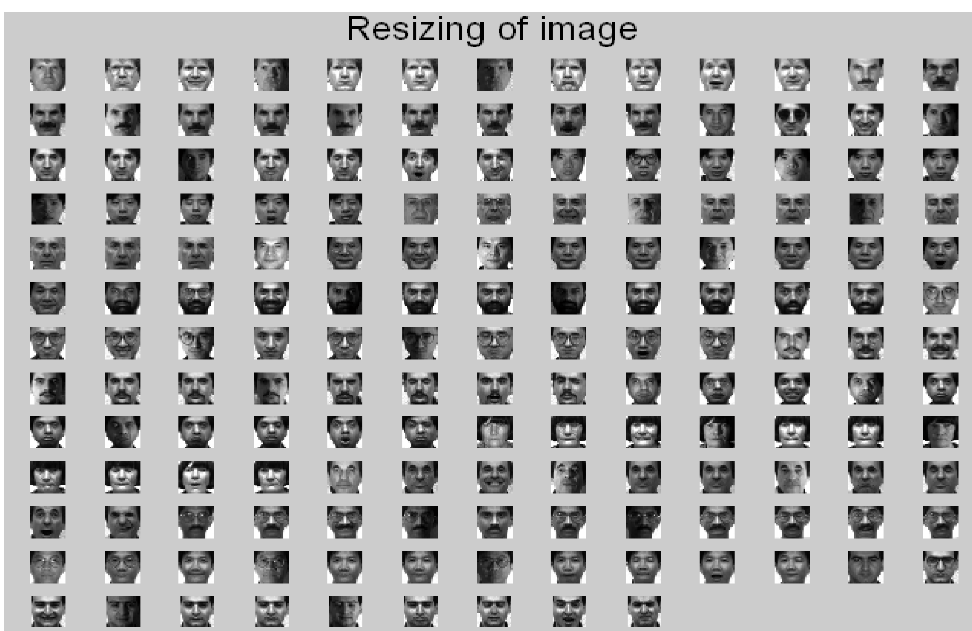


Figure 4.16 Resized Image Database

The Yale Face Database is used in order to test our method in the presence of headpose variations. The Yale Face Database contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad and sleepy, surprised, and wink. The whole database is shown in Figure 4.17.

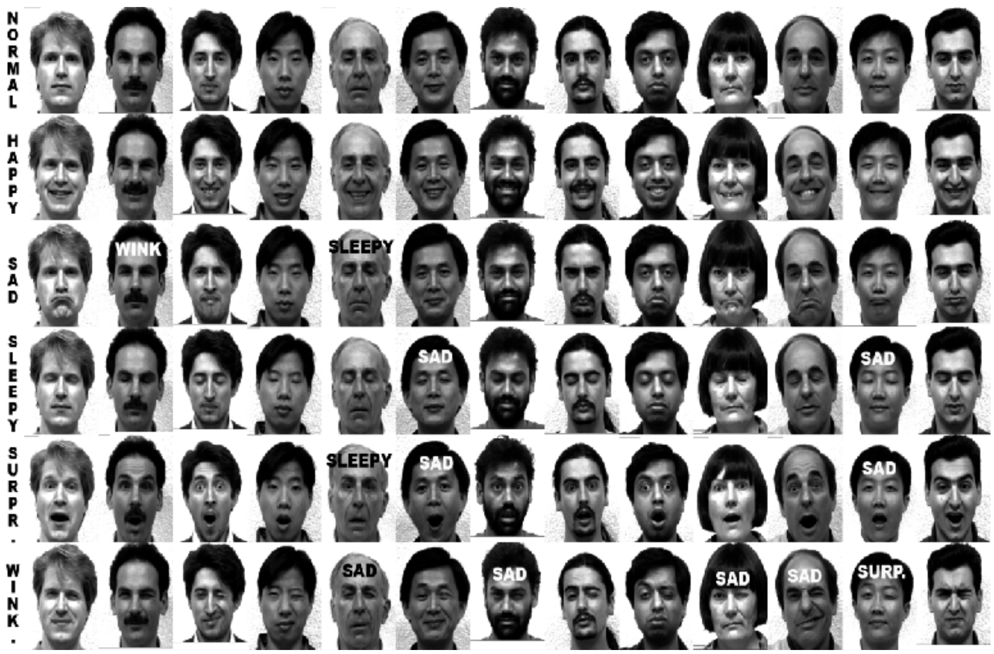


Figure 4.17 Whole Yale Face Database

Figure 4.18 Mean face for YALE face database

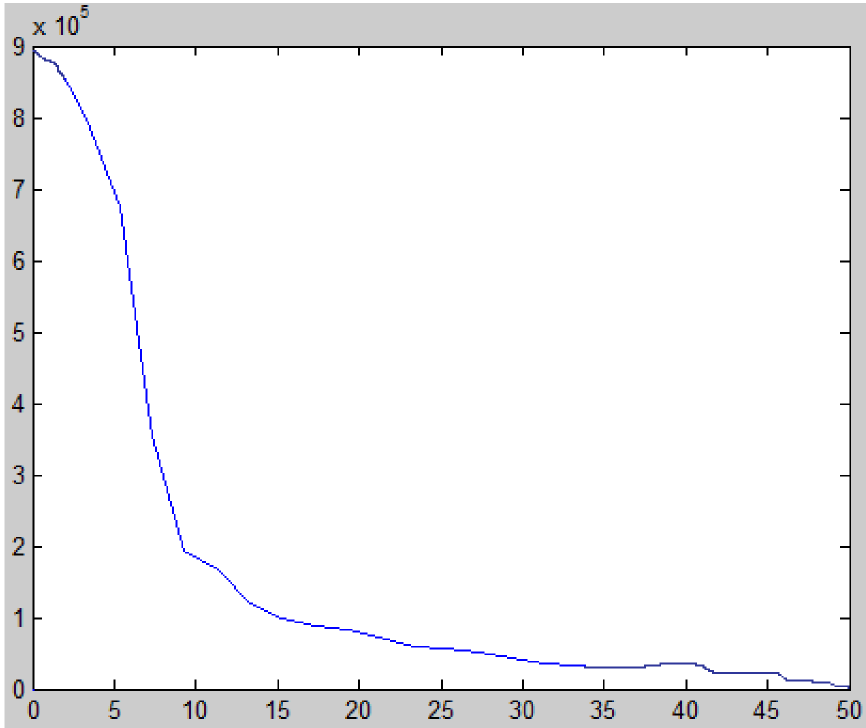


Figure 4.19 The Eigen values for YALE Face Database

For testing the whole database, the faces used in training, testing and recognition are changed and the recognition performance is given for whole database.

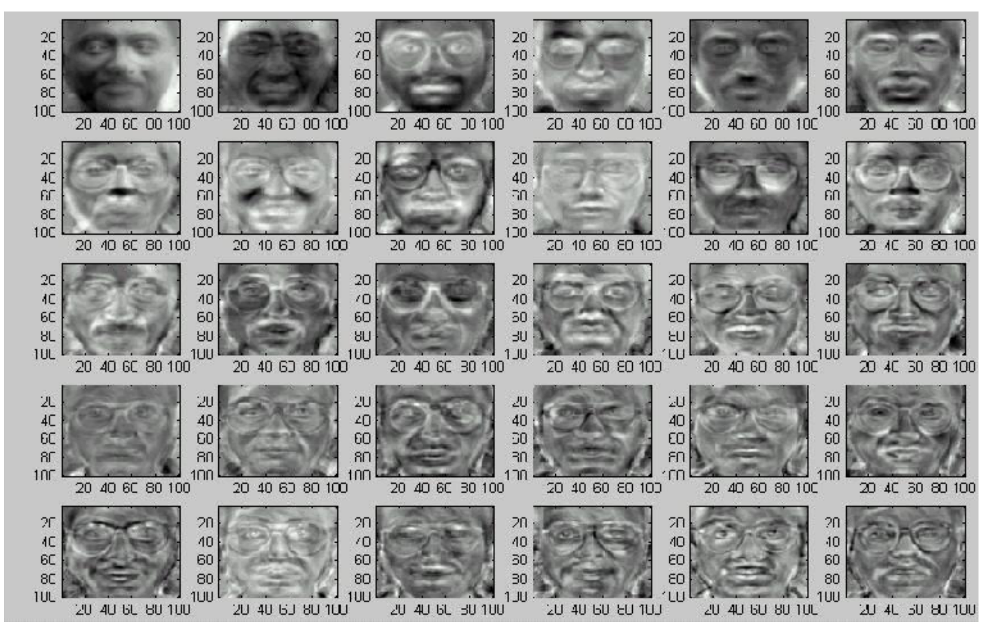


Figure 4.20 The top 30 eigen faces for the YALE Face Database

For this database, the mean face of the faces, the calculated top 30 (with highest eigenvalues) eigenfaces, and their corresponding eigenvalues are shown in Figure 4.18, Figure 4.19 and Figure 4.20 respectively.

4.1.4 Face Detection

1) Averaging Technique

Within a given database, all weight vectors of a like person are averaged together. This creates a "face class" where an even smaller weight matrix represents the general faces of the entire system. When a new image comes in, its weight vector is created by projecting it onto the face space. The face is then matched to the face class that minimizes the euclidean distance. A 'hit' is counted if the image matches correctly its own face class. A 'miss' occurs if the minimum distance matches to a face class of another person.

2) Removal Technique

This procedure varies only slightly from the averaging technique in one key way. The weight matrix represents the image projection vectors for images of the entire database. For empirical results, an image is removed from the system, and then projected onto the face space. The resulting weight vector is then compared to the weight vector of all images. The image is then matched to the face image that minimizes the euclidean distance. A 'hit' is counted if the tested image matches closest to another image of the same person. A 'miss' occurs when the image matches to any image of a different person. The main difference from the average technique is the number of possible images that the test face can match to that will still result in a hit.

Now from the Original YALE face database Figure 4.21 we fill detect the faces in the MATLAB R2007b.

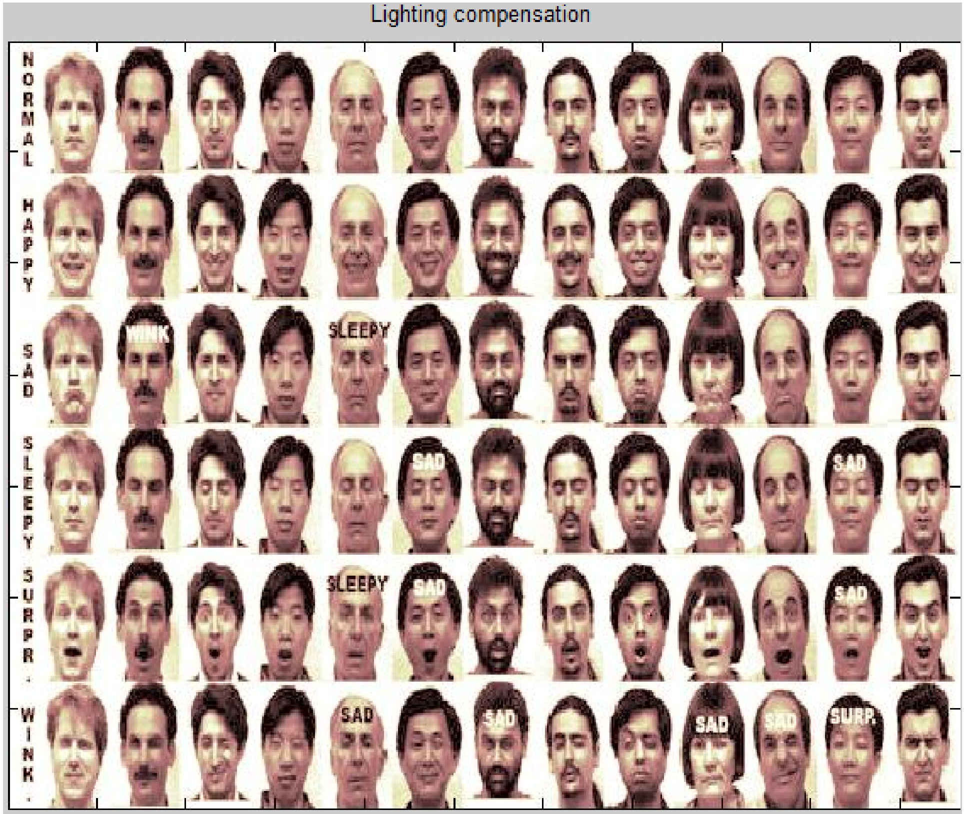


Figure 4.21 Lightning Compensation of the YALE Database

Figure 4.22 Skin Detection of the YALE Face Database

Figure 4.23 Noise removal from the skin detected

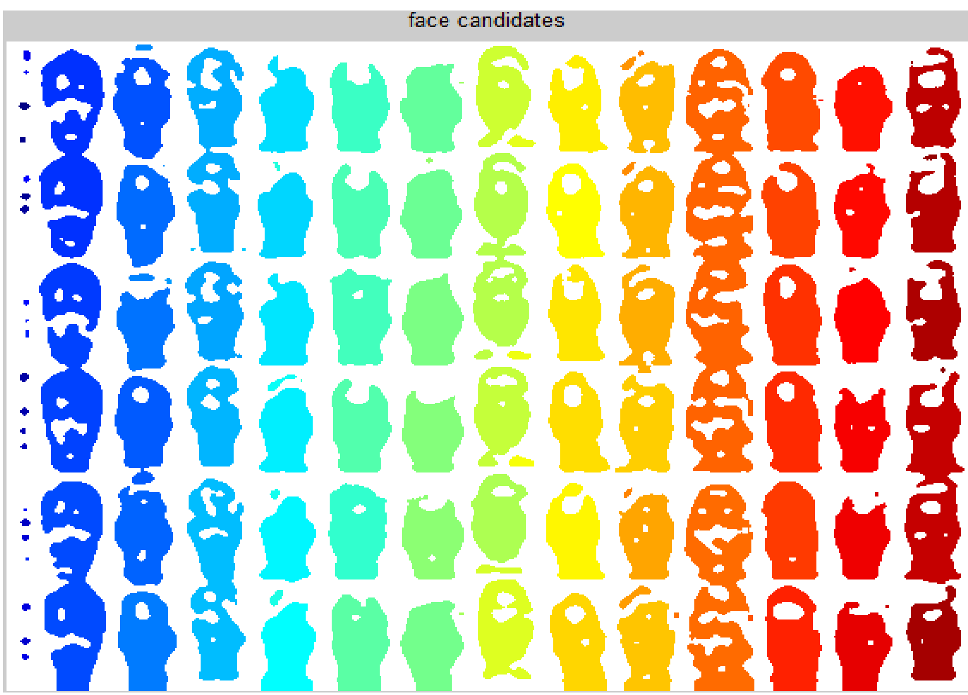


Figure 4.24 Faces that are detected.

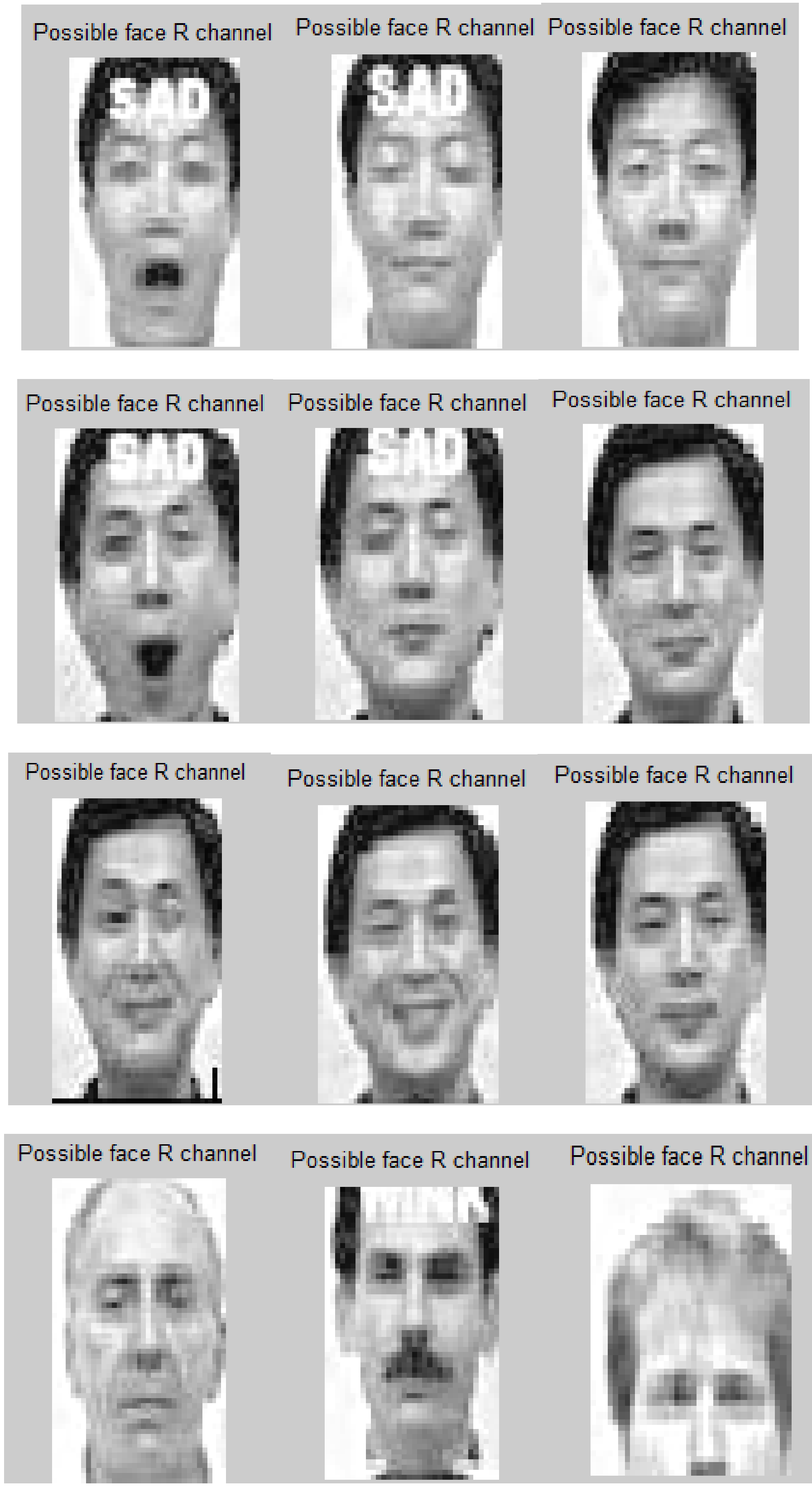


Figure 4.25 Possible faces that are detected.

Results that we obtained:

Number of faces found: 12

Indices of faces found: 12

CurBB =

19.5000	149.5000	36.0000	46.0000
64.5000	165.5000	34.0000	57.0000
188.5000	14.5000	35.0000	61.0000
229.5000	15.5000	36.0000	60.0000
229.5000	86.5000	36.0000	60.0000
229.5000	376.5000	37.0000	61.0000
230.5000	156.5000	36.0000	62.0000
230.5000	235.5000	36.0000	55.0000
230.5000	305.5000	36.0000	57.0000
481.5000	16.5000	34.0000	58.0000
481.5000	233.5000	33.0000	57.0000
481.5000	306.5000	33.0000	56.0000

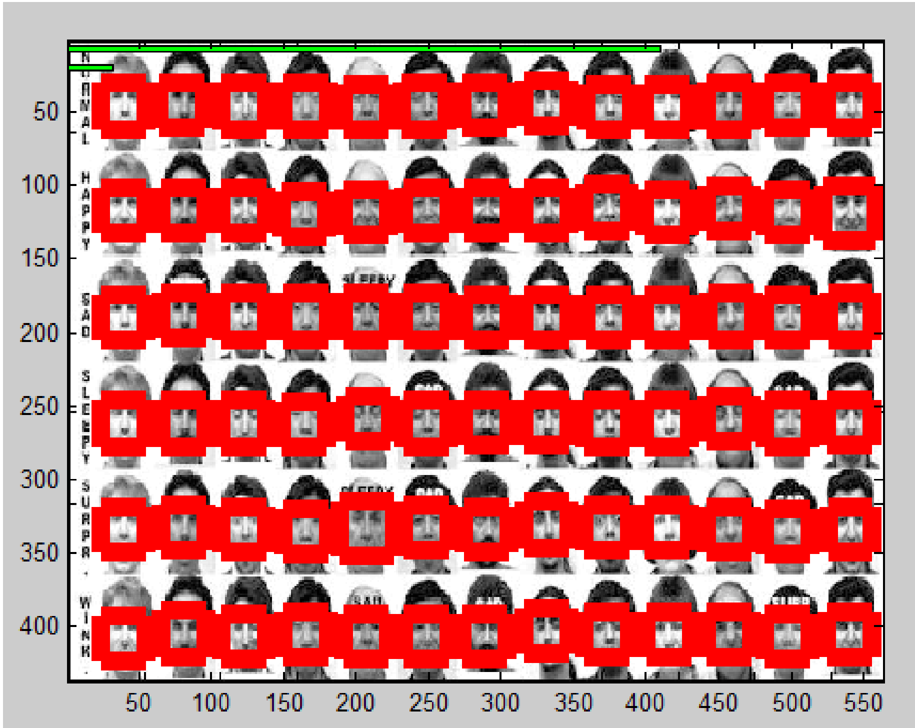


Figure 4.26 Faces detected from the Original YALE Face Database.

4.2 Face Recognition using the Neural Network with Back Propagation Algorithm.

Algorithm

1) Compute histogram and probabilities of each intensity level.
2) Set up initially $w_i(0)$ and $\mu_i(0)$. Where w_i are the class probabilities and μ_i is the class means.
3) Step through all possible thresholds t = 1……………maximum intensity.
4) Update w_i and μ_i.
5) Compute $\sigma_b^2(t)$ using the formula

$$\sigma_b^2(t) = \sigma^2 - \sigma_w^2(t) = w_1(t)w_2(t)[\mu_1(t) - \mu_2(t)]^2 \qquad (4.2.1)$$

6) Desired threshold correspond to the maximum $\sigma_b^2(t)$.

4.2.1 Thresholding

Thresholding is the operation of converting a grayscale image into a binary image. The key parameter in the thresholding process is the selection of the threshold value. All the pixels having values greater than threshold value are marked as "object" pixels and others as "background". There are mainly two types of thresholding techniques available

i). Global thresholding

ii). Local thresholding

The threshold level is calculated by the MATLAB command *level*.

Syntax

$$llevel = graythresh(J)evel = graythresh(J)$$

where, J is the image whose thresholding is to be done.

It computes a Global threshold (level) that can be used to convert an intensity image to a binary image with *im2bw*. Level is a normalized intensity value that lies in the range [0, 1].

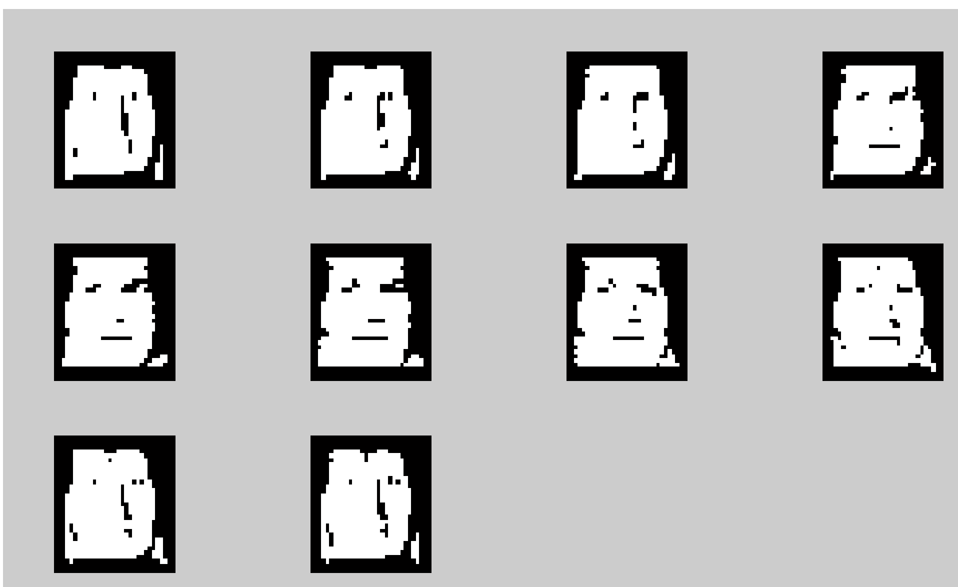


Figure 4.27 Thresholded images of typical face image

4.2.2 Classification Module

In this module, with the help of a pattern classifier, extracted features of the face image is compared with the ones stored in a face library (or face database). After doing this comparison, face image is classified as either known or unknown. Neural network is used as the classifier in present work which consists to two layers:

i). Input Layer
ii). Output Layer

4.2.3 Input Layer

This layer is also called the Gabor layer, receives the Gabor features. The number of nodes in this layer depends upon the size each image in Gabor features. The transfer function used for this layer is hyperbolic tangent sigmoid. To avoid variable number of nodes the size of image is fixed before. The number of nodes in present work is 961.

4.2.4 Output Layer

The number of nodes in output layer depends on the number of individual faces the network has to recognize. The Linear transfer function is used in this layer. If the person belongs to the stored database the classifier will return the value .9 corresponding to the node that belongs to that person. If the person does not belong to the database it will return a garbage value.

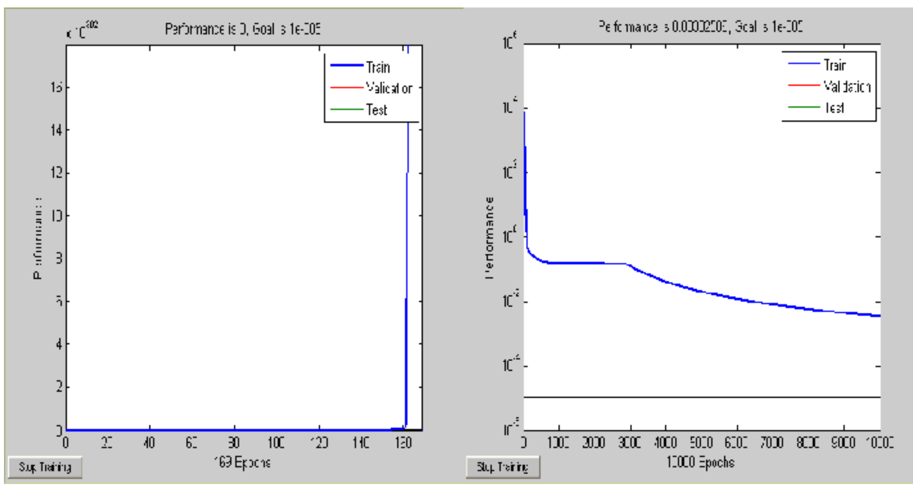
4.2.5 Training Set

The weights of the neural network are adjusted during the training. Batch training is used in present work to train the network. In batch training the weights and biases are updated after all the inputs and targets are presented. Feed forward back propagation network is created by the MATLAB command *newff*.

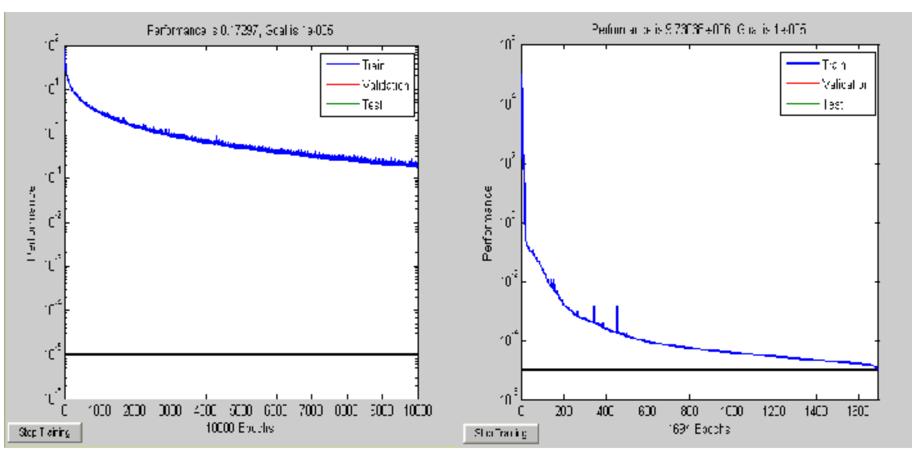
4.2.5.1 Choice of Training Algorithm

There are different back propagation algorithms for feedforward networks which use batch training are available in MATLAB. Some important ones are compared for their performance (time requirement and epochs) in the case of face recognition. In all the cases default values are considered with the goal $1e^{-005}$. In batch gradient descent (traingd) the weights and bias are updated in the direction of the negative gradient performance function while in batch gradient decent with momentum (traingdm), the momentum allows a network to respond not only to the local gradient, but also to recent trends in error surface. In adaptive learning rate algorithm (traingda) the learning rate is increased if the new errors are less than old errors and In Resilent back propagation training algorithm eliminates the harmful effects of the magnitude of the partial derivatives by just considering the sign of the derivative to determine the direction of weight update; the magnitude of the derivative has no effect on the weight update.

The performance of these algorithms is shown in figure 4.28



(a) (b)



(c) (d)

Figure 4.28 Training curve for (a) Gradient decent (b) Gradient decent with momentum (c) Adaptive rate (d) Resilent Back propagation algorithm.

It has been observed from the above figure that for gradient decent, gradient decent with momentum and adaptive rate algorithms the desired output is not achieved. Moreover epochs taken are also very large. Therefore, it is concluded from the above graphs that the Resilent back propagation learning algorithm is more efficient than all other above algorithms as the goal is achieved with lesser number of epochs and time required is also very less in this case. So the Resilent back propagation is used in present work for training the network.

Using the learning rate of 0.5 thirteen persons with six different poses per person has been taken for training the neural network and the network has been tested by considering five poses with different threshold values for the DWT which are used for training and remaining five poses per person taken from the data base. The results obtained are:

To evaluate the performance of the proposed method, we used the collection YALE face databases available in the public domain .There are 13 subjects with 6 images per subject for a total of 78 images.

Table 4.2 True test results

Total number of persons	Number of principal components	Neural Network Structure	Face Recognized (%)
13	1-28	28:20:78	93.37
13	1-36	36:30:78	93.89
13	1-50	50:40:78	94.78
13	***1-52***	***52:50:78***	***97.92***
13	1-64	64:60:78	95.76

As we see in the Table 4.1 shows that for the YALE database best results have been obtained when 52 components from PCA have been selected and 50 neurons in the hidden layer have been used.

Chapter 5

Conclusion

The face recognition system using DWT, PCA and neural network has been implemented using MATLAB version 7.5.0.267 (R2007b). The image is first preprocessed (DWT, wavelet maxima, eigen faces, resized, and thresholded) and DWT are used to extract the face feature like eyes, mouth, nose etc. to make fit for to be given to the neural network. The training algorithm is chosen as Resilent back propagation by considering the number of epochs and time taken to obtain the final output goal $1e^{-005}$. The choice of learning rate is very critical parameter in the training process. The experiments were conducted to obtain the optimum learning rate which comes out to be 0.5 in this case considering the final goal and the success rate. The Yale database has been used in the present work and the experiment conducted shows the 97.92% recognition rate with 50 neurons in the hidden layer.

The algorithm is also implemented on the colored images for recognition and the processing time is also reduced as the system is first detecting the face images from the database.

Future Scope

A face recognition system using PCA, DWT and neural network has been developed. Although a lot of effort has been made for the development of the system, still following points may be considered for the improvement of the system in future.

i) The present work has been implemented with the face images in frontal view it can be extended with the side view faces as well.

ii) Algorithm can be improved for recognizing the faces more effectively that is partially occluded by the objects.

iii) In the present work a linear technique of Principal Component Analysis has been used other linear/non-linear techniques may be tried to improve the recognition rate of the system.

REFRENCES

[1] Gupta, S., Singla, S.K., "A Face identification system using neural network". *International Journal of Information Technology & Knowledge Management.* 2008, pp. 227-234.

[2] http://en.wikipedia.org/wiki/Face.

[3] http://en.wikipedia.org/wiki/Fusiform_face_area

[4] Matthew TURK, "A Random Walk through Eigenspace". *IEICE TRANS. INF. & SYST.*, VOL.E84–D, NO.12 DECEMBER 2001, pp. 1586-1595.

[5] Gong, S., McKenna, S., Psarrou, A., "DYNAMIC VISION From Images to Face Recognition", pp. 5-6.

[6] IAPR. *International Conference on Audio and Video-Based Person Authentication,* Lecture Notes in Computer Science, Crans-Montana, Switzerland, March 1997. Elsevier Science.

[7] IAPR. *International Conference on Audio and Video-Based Person Authentication,* Lecture Notes in Computer Science, Washington DC, USA, March 1999. Elsevier Science.

[8] IEEE. *International Conference on Face* & *Gesture Recognition,* Zurich, Switzerland, March 1995. IEEE Computer Society.

[9] IEEE. *International Conference on Face* & *Gesture Recognition,* Killington, Vermont, USA, October 1996. IEEE Computer Society.

[10] IEEE. *International Conference on Face* & *Gesture Recognition,* Nara, Japan, April 1998. IEEE Computer Society.

[11] E.S. Spelke, G. Gutheil, and G. Van de Walle. The development of object perception. In S.M. Kosslyn and D.N. Osherson, editors, *Visual Cognition*, volume **2** of *An Invitation to Cognitive Science,* MIT Press, Cambridge, Massachusetts, 1995, pp. 297-330.

[12] S. Grossberg. "How does the cerebral cortex work? Learning, attention and grouping by the laminar circuits of visual cortex." Technical report, Department of Cognitive and Neural Systems, University of Boston, Boston, USA, April 1998.

[13] S. Grossberg and G.O. Stone, "Neural dynamics of word recognition and recall: attention priming, learning and resonance." *Psychological Review,*

1986, pp. 46-74.

[14] H.H. Nagel, "From image sequences towards conceptual descriptions." *Image and Vision Computing*, 1988, pp. 59-74.

[15] B. Neumann, "Natural language description of time varying scenes." In *Semantic Structures.* Lawrence Erlbaum Associates, 1989, pp. 167-206.

[16] J. J. Weng and D. L. Swets, "Face Recognition", in A. K. Jain, R. Bolle, and S. Pankanti (Editors), *Biometrics: Personal Identification in Networked Society*, Kluwer Academic Press,1999.

[17] Hietmeyer, R. Biometric identification promises fast and secure processing of airline passengers. *The International Civil Aviation Organization Journal*, 55(9), 2000, pp. 10–11.

[18] Machine Readable Travel Documents (MRTD). http://www.icao.int/mrtd/overview/overview.cfm.

[19] Zhao, W., Chellappa, R., Phillips, P.J. and Rosenfled, A. "Face Recognition: A Literature Survey." *ACM Computing Surveys*, Vol 35, No. 4, 2003, pp. 399-458.

[20] Zhao, W., Chellapa, R., Rosenfeld, A., and Phillips, P.J., "Face Recognition: A Literature Survey," *Technical Report, Univ. of Maryland*, 2000. pp. 12-14

[21] Temdee, P., Khawparisuth, D. and Chamnongthai, K., "Face Recognition by using Fractel Encoding and Back propagation Neural Network", *Fifth International Symposium on Signal Processing and its Applications*, ISSPA'99, Australia, Aug.1999, pp.159-161.

[22] Bruce, V., "Identification of Human Faces", *Seventh International Conference on Image Processing and its Applications*, vol.2, July1999 pp.615-619.

[23] Li, S. Z. and Jain, Anil K., editors. *Handbook of Face Recognition*. Springer-Verlag, 2005, pp. 10-11.

[24] Phillips, P.J., Grother, P., Micheals, R.J., Blackburn, D.M., Tabassi, E., and Bone, J .M. FRVT2002: *Overview and Summary.* Available at http://www.frvt.org/FRVT2002, March 2003.

[25] Jain, A.K., Stan, Z., "Handbook of Face Recognition," Springer, 2004, pp. 21-22.

[26] R. Chellappa, C. L. Wilson and S. Sirohey, "Human and Machine Recognition of Faces: A Survey", *Proceedings of the IEEE*, Vol. 83,

No. 5, May 1995.

[27] P. Temdee, D. Khawparisuth, and K. Chamnongthai, “Face Recognition by Using Fractal Encoding and Backpropagation Neural Network”, pp. 159-161, *5th International Symposium on Signal Processing and its Applications*, ISSPA ’99, Australia, August 1999.

[28] J. Huang, “Detection Strategies For Face Recognition Using Learning and Evolution”, PhD. Thesis, George Mason University, May 1998.

[29] Zhao, W., Chellappa, R., Rosenfeld, A., and Phillips, P.J. Face Recognition: A Literature Survey. *ACM Computing Surveys*, 2003, pp. 399–458.

[30] Singh, R., Vatsa, M., Ross, A. and Noore, A. Performance Enhancement of 2D Face Recognition via Mosaicing. In *Proceedings of 4th IEEE Workshop on Automatic Identification Advanced Technologies (AutoID)*, 2005, pp.63–68.

[31] Liu, X. and Chen, T. Pose Robust Face Recognition Based on Mosaicing –An Example Usage of Face In Action (FIA) Database. In *Demo session of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2004, pp. 154-160.

[32] Wiskott, L., Fellous, J.M., Kr¨uger, N. and Malsburg, C. von der. Face Recognition by Elastic Bunch Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 1997, pp.775–779.

[33] Turk, M. and Pentland, A. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1), 1991, pp.71–86.

[34] Duda, R. O., Hart, P. E. and Stork, D. G. *Pattern Classification*. Wiley-Interscience Publication, 2000, pp. 54-56.

[35] Belhumeur, P., Hespanha, J. and Kriegman, D. Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 1997, pp.711–720.

[36] Yu, H and Yang, J. A direct LDA algorithm for high-dimensional data with application to face recognition. *Pattern Recognition*, 34(10), 2001, pp. 2067–2070.

[37] Lu, J., Plataniotis, K.N. and Venetsanopoulos, A.N. Face recognition using LDA-based algorithms. *IEEE Transactions on Neural Networks*, 14(1), 2003, pp, 195–200.

[38] Zheng, W., Zou, C. and L. Zhao, L. Real-Time Face Recognition Using Gram-Schmidt Orthgonalization for LDA. In *IEEE Conference of Pattern*

Recognition, 2004, pp.403–406.

[39] Bartlett, M. S., Movellan, J. R. and Sejnowski, T.J. Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, 13(6), 2002, pp.1450–1464.

[40] Penev, P.S. and Atick, J.J. Local Feature Analysis: A General Statistical Theory for Object Representation. *Network: Computation in Neural Systems*, 7(3), 1996, pp. 477–500.

[41] Phillips, P. J ., Grother, P., Micheals, R.J., Blackburn, D.M., Tabassi, E. and Bone, M. Face Recognition Vendor Test 2002. Evaluation report, National Institute of Standards and Technology, 2003, pp. 302-309.

[42] Lawrence, S., Giles, C. L., Tsoi, A. C. and Back, A. D. Face Recognition: A Convolutional Neural-Network Approach. *IEEE Transactions on Neural Networks*, 8(1), 1997, pp. 98–113.

[43] Gutta, S., Huang, J. R. J. , Jonathon, P . and Wechsler, H . Mixture of experts for classification of gender, ethnic origin, and pose of human faces. *IEEE Transactions on Neural Networks*, 11(4), 2000, pp. 948–960.

[44] Zhang, D., Peng, H., Zhou, J. and Pal, S. K. A novel face recognition system using hybrid neural and dual eigenspaces methods. *IEEE Transactions on Systems, Man and Cybernetics - Part A*, 32(6), 2002, pp. 787–793.

[45] Vapnik, V. N. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995, pp. 45-46.

[46] P. Temdee, D. Khawparisuth, and K. Chamnongthai, "Face Recognition by Using Fractal Encoding and Backpropagation Neural Network", Fifth International Symposium on Signal Processing and its Applications, ISSPA '99, Brisbane, Australia, August 1999, pp. 159-161.

[47] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. Von der Malsburg, R. P. Wurtz, and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture, IEEE Transactions on Computers, Vol. 42, March 1993", pp. 300-310.

[48] Heisele, B., Ho, P. and Poggio, T. Face Recognition with Support Vector Machines: Global versus Component-based Approach. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2, 2001, pp. 688–694.

[49] Baudat, G. and Anouar, F. Generalized Discriminant Analysis Using a Kernel Approach. *Neural Computation*, 12(10), 2000, pp. 2385–2404.

[50] Yang, M. H. Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition using Kernel Methods. In *IEEE Conference on Automatic Face and Gesture Recognition*, 2002, pp. 215–220.

[51] Muller, K. R., Tsuda, K., Mika, S., Ratsch, G. and Scholkopf, B. Introduction to Kernel-Based Learning Algorithms. *IEEE Transactions on Neural Network*, 12(2), 2001, pp. 181–202.

[52] Lu, J., Plataniotis, K. N. and Venetsanopoulos, A. N. Face recognition using kernel direct discriminant analysis algorithms. *IEEE Transactions on Neural Networks*, 14(1), 2003, pp. 117–126.

[53] Mika, S., Ratsch, G., Weston, J., Scholkopf, B. and Muller, K. R. Fisher Discriminant Analysis with Kernels. In *Proceedings of IEEE Neural Networks for Signal Processing Workshop*, 1999, pp. 215-221.

[54] Kim, K. I., Jung, K. and Kim, H. J. Face recognition using kernel principal component analysis. *IEEE Signal Processing Letters*, 9(2), 2002, pp. 40–42.

[55] Bach, R. and Jordan, M. I. Kernel Independent Component Analysis. *Journal of Machine Learning Research*, 3, 2002, pp. 1–48.

[56] Vasilescu, M. A. O. and Terzopoulos, D. Multilinear analysis of image ensembles: TensorFaces. In *Proceedings European Conference on Computer Vision*, 2002, pp. 447–460.

[57] Lathauwer, L. D., Moor, B. D. and Vandewalle, J. A multilinear singular value decomposition. *SIAM Journal of Matrix Analysis and Applications*, 21(4), 2000, pp. 1253–1278.

[58] Roweis, S. and Saul, L. K. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500), 2000, pp. 2323–2326.

[59] He, X. and Niyogi, O. Locality preserving projections. *Advances in Neural Information Processing Systems*, 16, 2003, pp. 109-117.

[60] Tenenbaum, J. B., Silva, V. and Langford, J. C. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(12), 2003, pp. 2319–2323.

[61] Belkin, M. and Niyogi, P. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. *Advances in Neural Information Processing*

Systems 14, MIT Press, Cambridge, 2002, pp. 585–591.

[62] He, X., Yan, S., Hu, Y., Niyogi, P. and Zhang, H. Face Recognition Using Laplacian faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3), March 2005, pp. 617-624.

[63] S.N. Sivanandam, S. Sumathi and S.N. Deepa, “Introduction to Neural Networks Using MATLAB 6.0,” TATA McGraw Hill Publishing Company Limited, 2006.