Improved Face Recognition using Image Resolution Reduction and Optimization of Feature Vector

By

Muhammad Almas Anjum

A thesis Submitted to the College of Electrical and Mechanical Engineering
National University of Sciences and Technology Rawalpindi Pakistan,
in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Supervisor: Dr Muhammad Younus Javed

Department of Computer Engineering
College of Electrical and Mechanical Engineering
National University of Sciences and Technology (NUST) Rawalpindi Pakistan,

January, 2008

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

First of all I am sincerely grateful to Allah almighty for blessing me courage, audacity and strength for completion of this research work. I would like to express my deep gratitude to my advisor Dr Muhammad Younus Javed for his continuous encouragement, advice, help and invaluable suggestions. He is such a nice, generous, helpful and kindhearted person that I feel really happy, comfortable and unconstrained with him during my PhD study. I owe my research achievements to his experienced supervision. Many thanks are also due to my PhD Guidance and Examination Committee (GEC) which includes Dr Muhammad Ejaz, Dr Muhammad Shoib. A. Khan and Dr Muid Mufti for their valuable ideas, guidance, suggestions, and encouragement. I also acknowledge the help and guidance extended by Dr Hamid R. Arabnia during my research work.

I would like to thank all my friends Mr Javed Akhtar, Mr Salman Ahmed, Abdul Basit, Mr Shams who have helped me during my Ph.D. study at Department of Computer Engineering College of E&ME (NUST). Special thanks are due to Dr Shaleeza supporting and encouraging me generously once it was really required.

This thesis could not have been completed without the endless support of my brothers, sisters and in particular my father and mother who always remained a source of encouragement for me during my complete duration of studies. I can not put out of my mind how my parents have been praying to Allah almighty for my success and health at all critical moments of my PhD work. This always remained a main instrument and motive for me to stay committed to my goal of research. I must admire the constant support of my younger sisters Kalsoom and Rukhsana for pursuing my PhD.

A great deal of thanks goes for my wife Sabeen and little daughters Aamina and Aleena for their endless support and encouragement even during the most demanding phases of this research. I must acknowledge their scarifies of time which they owed to me for my PhD research work.

This research wok is dedicated to

My father and mother

My Brothers and sisters

My wife and daughters Aamina & Aleena

Abstract

Face recognition is a difficult problem that involves automated matching of a given face image with corresponding person's image(s) in a database. Face recognition finds application in areas like surveillance & security, digital libraries and human computer interactions. Successful, speedy and practically feasible face recognition method depends heavily on the choice of feature vector used for classification and addressing the curse of image dimension. The dimension reduction and the skill to acquire minimum size of feature vector required for face recognition for diverse facial expressions is a challenging task in face recognition. Dimension reduction results in removal of irrelevant variables alongwith noise therein and a lower computation complexity of subsequent processing.

This dissertation addresses the challenges of dimension reduction, choice of minimum size feature vector for face recognition and minimization of adverse effects of varying facial expressions on the recognition through reduction in image resolution. In preprocessing of face images, scale normalization is carried out through a novel scale normalization algorithm to retain only the facial part of images. This helps in reducing computational complexity by restricting dimensions of image to face region only. Tilt of face images is removed by calculating the gradient between the two eyes and applying the reverse rotation. The issue of dimensionality is addressed first by gradually reducing image resolution through spatial domain low pass filtering followed by decimation. The second method involves novel coefficient selection strategies to choose the minimum dimension of feature vector required for recognition with maximum recognition rate and reduced computational complexity. Face images with varying image resolution are obtained by varying the decimation factor. The effects of variation in image resolution on face recognition have been evaluated using template matching and Principle Components Analysis (PCA) based face recognition techniques. Classical PCA technique has been modified into sub-holistic PCA. Better recognition rate is achieved using modified PCA method with reduced image resolution.

Improved recognition rate results are reported using novel coefficients selection and optimization methods in Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT) and Discrete wavelets Transform (DWT) based face recognition

methods. The experiments are carried out for various image resolutions using five different datasets. Improved recognition rate of 97.2% (template matching), 87% (PCA), 94% (Sub-holistic PCA), 100% (DFT), 95.75% (DCT) and 99.25% (DWT) is achieved at a specific image resolution for different datasets.

The resolution reduction method used with square images is then extended to hexagonal images. A new technique based on Diagonal grow and Butterfly structure methodology has been developed for sampling and indexing hexagonal structure in hexagonal image processing frame work. Proposed strategy offer less pixel redundancy as compared to existing techniques. Reduction in pixel redundancy varies according to size of square image.

CHAPTER 1

INTRODUCTION

1.1 Introduction

A pattern is the description of an object. A human being is a very sophisticated information system, partly because it possesses a superior pattern recognition capability. Pattern recognition can be defined as the categorization of input data into identifiable classes via the extraction of significant features or attributes of the data from a background of irrelevant details. Pattern recognition is the study of how machines can observe the environment, learn to distinguish patterns of interest from their background and make sound and reasonable decisions about the categories of the patterns. In spite of almost 55 years of research, design of a general purpose machine pattern recognizer remains an elusive goal. Automatic recognition, description, classification and grouping of patterns are important problems in a variety of engineering and scientific disciplines such as biology, psychology, medicine, marketing, computer vision, artificial intelligence, and remote sensing. Watanabe [1] defines a pattern as opposite of a chaos; it is an entity, vaguely defined, that could be given a name. For example, a pattern could be a fingerprint image, a handwritten cursive word, a human face, or a speech signal etc. Given a pattern, its recognition/classification may consist of one of the following two tasks:

1.1.1 Supervised Classification

In this approach, the input pattern is identified as a member of a predefined class.

1.1.2 Unsupervised Classification

The pattern is assigned to a hitherto unknown class. The recognition problem here is being posed as a classification or categorization task, where the classes are either defined by the system designer (in supervised classification) or learnt based on the similarity of patterns.

1.2 Pattern recognition has been renewed recently due to emerging applications as shown in Table 1.1. Such applications are not only challenging but also computationally more demanding. These applications include data mining (identifying a pattern, e.g., correlation, or an outlier in millions of multidimensional patterns), document classification (efficiently searching text documents), financial forecasting, organization and retrieval of multimedia databases and biometrics (personal identification based on various physical attributes such as face and fingerprints). Picard [2] has identified a novel application of pattern recognition called effective computing which will give a computer the ability to recognize and express emotions, to respond intelligently to human emotion and to employ mechanisms of emotion that contribute to rational decision making. The design of a pattern recognition system essentially involves the following three aspects:

- Data acquisition and preprocessing
- Data representation
- Decision making

Table 1.1 Pattern recognition applications

Domain	**Application**	**Pattern**
Bioinformatics	Sequence analysis	DNA/Protein Sequence
Data Mining	Searching for meaningful patterns	Points in multidimensional space
Documents Classification	Internet search	Text documents
Documents Image Analysis	Reading machine for the blinds	Document image
Industrial Automation	Printed circuit board inspection	Intensity or range image
Multimedia Database Retrieval	Internet search	Video clip
Biometric Recognition	Personal identification	Face, iris, fingerprint etc
Remote Sensing	Forecasting crop yield	Multispectral image
Speech Recognition	Telephone directory	Speech waveform

It is generally agreed that a well-defined and sufficiently constrained recognition problem (small intraclass variations and large interclass variations) will lead to a compact pattern representation and a simple decision making strategy. Learning from a set of examples (training set) is an important and desired attribute of most pattern recognition systems. The four best known approaches for pattern recognition are:

- Template matching
- Statistical classification
- Syntactic or structural matching
- Neural networks

1.2.1 Template Matching

One of the simplest and earliest approach to pattern recognition is based on template matching, where a template (typically, a 2D shape) or a prototype of the pattern to be recognized is available. The pattern to be recognized is matched against the stored template while taking into account all allowable pose (translation and rotation) and scale changes. The similarity measure, often a correlation, may be optimized based on the available training set. Often, the template itself is learnt from the training set. Template matching is computationally demanding, but the availability of faster processors has now made this approach more feasible. The rigid template matching (although effective in some application domains) has a number of disadvantages. For instance, it would fail if the patterns are distorted due to the imaging process, viewpoint change, or large intraclass variations among the patterns. Deformable template models [3] or rubber sheet deformations [4] can be used to match patterns when the deformation cannot be easily explained or modeled directly.

1.2.2 Statistical Classification

In the statistical approach, each pattern is represented in terms of d-features or measurements and is viewed as a point in a d-dimensional space. The goal is to choose those features that allow pattern vectors belonging to different categories to occupy compact and disjoint regions in a d-dimensional feature space. The effectiveness of the representation space (feature set) is determined by how well patterns from different classes can be separated. Given a set of training patterns from each class, the objective is to establish decision boundaries in the feature space which separate patterns belonging to different classes. The decision boundaries are determined by the probability distributions of the patterns belonging to each class, which must either be specified or learnt [5,6]. One can also take a discriminant analysis-based approach to classification.

1.2.3 Syntactic or Structural Matching

In many recognition problems involving complex patterns, it is more appropriate to adopt a hierarchical perspective where a pattern is viewed as being composed of simple subpatterns which are themselves built from yet simpler subpatterns [7]. The simplest/elementary subpatterns to be recognized are called primitives and the given complex pattern is represented in terms of the interrelationships between these primitives. In syntactic pattern recognition, a formal analogy is drawn between the structure of patterns and the syntax of a language. The patterns are viewed as sentences belonging to a language, primitives are viewed as the alphabet of the language, and the sentences are generated according to a grammar. Thus, a large collection of complex patterns can be described by a small number of primitives and grammatical rules. The grammar for each pattern class must be inferred from the available training samples.

1.2.4 Neural Networks

Neural networks can be viewed as massively parallel computing systems consisting of an extremely large number of simple processors with many interconnections. Neural network models attempt to use some organizational principles (such as learning, generalization, adaptivness, fault tolerance and distributed representation, and computation) in a network of weighted directed graphs in which the nodes are artificial neurons and directed edges (with weights) are connections between neuron outputs and neuron inputs. The main characteristics of neural networks are that they have the ability to learn complex nonlinear input-output relationships, use sequential training procedures, and adapt themselves to the data.

1.3 Dimensionality Reduction in Pattern Recognition Models

There are two main reasons to keep the dimensionality of the pattern representation (i.e., the number of features) as small as possible: measurement cost and classification accuracy. A limited, yet salient, feature set simplifies both the pattern representation and the classifiers that are built on the selected representation. Consequently, the resulting classifier will be faster and will use lesser memory. Few feature extraction and projection methods in relation to dimension reduction are shown in Table 1.2.

Table 1.2. Feature Extraction and Projection Methods

Method	**Property**	**Comments**
Principal Component Analysis (PCA)	Linear, fast and eigenvector-based	Traditional, eigenvector based method, also known as Karhunen-Loeve expansion and good for Gaussian data
Linear Discriminant Analysis	Supervised linear map;eigenvector-based	Better than PCA for classification
Projection Pursuit	Linear, iterative and non –Gaussian	Mainly used for interactive exploratory data-analysis
Independent Component Analysis (ICA)	Linear iterative and non-Gaussian	Blind source separation used for de-mixing non-Gaussian distributed source
Kernel PCA	Nonlinear and eigenvector-based	PCA-based method, using a kernel to replace inner products of pattern vector
PCA Network	Linear and Iterative	Auto-associative neural network with linear transfer function and just one hidden layer
Non linear Auto-associative Network	Nonlinear map; non Gaussain and iterative method	Bottleneck network with several hidden layers; the non linear map is optimized by a nonlinear reconstruction
Multidimensional Scaling (MDS), and Sammon's Projection	Nonlinear and iterative	Often poor generalization, sample size limited ; noise sensitive mainly used for 2-dimensional visualization.
Self –Organization Map (SOM)	Nonlinear and iterative	Based on a grid of neurons in the feature space, suitable for extracting space of low dimensionality

1.4 Classifiers

Once a feature selection or classification procedure finds a proper representation, a classifier can be designed or the nearest mean classifier can be viewed as finding the nearest subspace [8]. The second main concept used for designing pattern classifiers is based on the probabilistic approach. Few classifiers are shown in Table 1.3.

Table 1.3 Classification Methods

Method	**Property**	**Comments**
Template matching	Assign patterns to the most similar template	The templates and the metric have to be supplied by the user, the procedure may include nonlinear normalization scale (metric) dependent
Nearest mean classification	Assign patterns to the nearest class mean	Almost no training needed, fast testing scale(metric)
Near neighbor rule	Assign patterns to the class of the nearest training patterns	No training, robust performance, slow testing scale (metric) dependent
Bayes plug-in	Assign pattern to the class which has the maximum estimate posterior probability	Yield simple classification (linear or quadratic) for Gaussian distribution and sensitive to density estimation error
Logistic classifier	Maximum likelihood rule for logistic posterior probabilities	Linear classifier iterative, procedures optimal for a family of different distributions (Gaussian), suitable for mix data types
Fisher linear discriminant	Linear classification using Mean Square Error (MSE) optimization	Simple and fast; similar to Bayer plug in
Binary decision tree	Find a set of threshold for a pattern dependent sequence of feature	Iterative training procedure, over training sensitive, needs pruning fast testing
Multi – layer perceptron (feed – forward neural network)	Iterative MSE optimization of two or more layers of perceptrons	Sensitive to training parameters, slow training and nonlinear classification function
Support vector classifier	Maximize the margin between the classes by selecting a minimum number of support vectors	Scale dependant, iterative, slow training method

1.5 Fundamental Problems in Pattern Recognition System Design

The design of an automatic pattern recognition system generally involves several major problem areas which include:

First of all, one has to deal with the representation of input data which can be measured from the objects to be recognized. This is the sensing problem. Each measured quantity describes a characteristic of the pattern or object. The set of patterns belonging to the same class corresponds to an ensemble of points scattered within some region of the measurement space. A simple example of this is shown in Figure.1.1 for two pattern classes denoted by Z_1 *and* Z_2. Each pattern is characterized by two measurements i,e. length and weight. The pattern vector, therefore, is in the form of $x = [x_1 \ \ x_2]^T$.

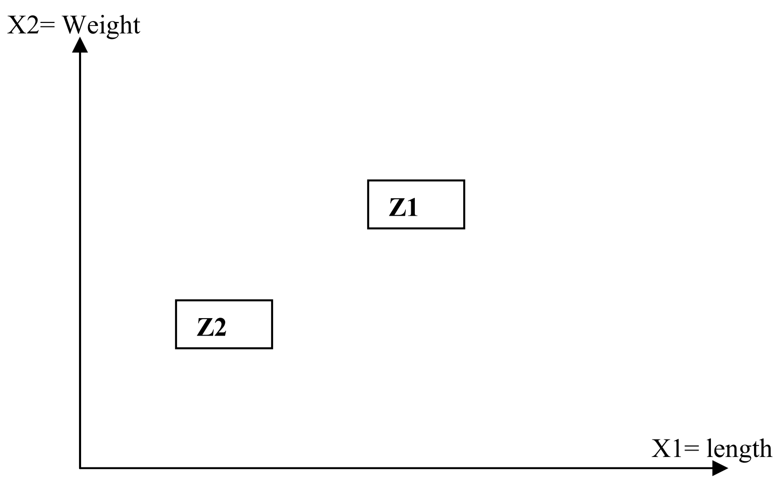


Figure 1.1. Two disjoint pattern classes

The second problem in pattern recognition concerns the extraction of characteristic features or attributes from the received input data and the reduction of the dimensionality of pattern vectors. This is often referred to as the pre-processing and the feature extraction problem.

The third problem in pattern recognition system design involves the determination of the optimum decision procedures, which are needed in the identification and classification process. After the observed data from patterns to be recognized have been expressed in the form of pattern points or measurement vectors in the pattern space, we want the machine to decide to which pattern class these data belong.

1.6 Biometrics

Pattern recognition encompasses a wide variety of applications and one of the current and emerging application is biometrics. It operates by acquiring biometric data from an individual, extracting a feature set from the acquired data, and comparing this feature set against the template set in the database. Traditionally passwords (knowledge-based security) and ID cards (token-based security) have been used to restrict access to secure systems. However, security can be easily breached in these systems when a password is divulged to an unauthorized user or a card is stolen by an impostor. The emergence of biometrics has addressed the problems that plague traditional verification methods. Biometrics refers to the automatic identification (or verification) of an individual (or a claimed identity) by using certain physiological or behavioral traits associated with the person. Few examples of biometrics are shown in Figure 1.2.

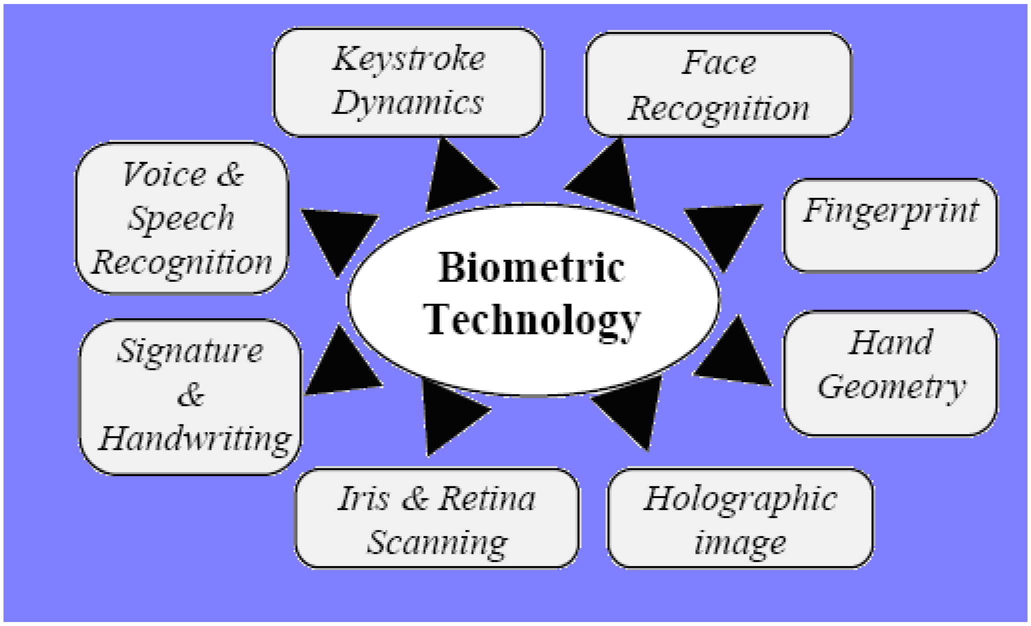


Figure 1.2. Examples of Biometrics

Biometrics means face recognition, voice and speech recognition, dynamic signature and handwriting capture, eyes (iris and retinal) identification, hand geometry, fingerprint (palm print) identification and keystroke dynamics. Humans have used these body characteristics for thousands of years to recognize each other. In a number of cases, thermal imaging of living objects is also used. Possibly, it should be classified as a separate direction of biometrics. Biometrics imposes a number of requirements for choice and use of biometrics methods: (i) whether it meets the requirements of a particular application (for example, for super smart cards, the signature and face identification are considered as good combination), (ii) whether it is possible to realize the chosen biometrics by using identical mathematical and algorithmic methods (for example, signature and face identification methods have little in common).

While biometric systems have their limitations, they have an edge over traditional security methods in that they cannot be easily stolen or shared. Besides bolstering security, biometric systems also enhance user convenience by alleviating the need to design and remember passwords. Moreover, biometrics is one of the few techniques that can be used for negative recognition where the system determines whether the person is who he or she denies to be.

1.7 Biometrics Qualifying Requirements

Any human physiological and/or behavioral characteristic can be used as a biometric characteristic as long as it satisfies the following requirements:

- *Universality*: Each person should have the characteristic.
- *Distinctiveness*: Any two persons should be sufficiently different in terms of the characteristic.
- *Permanence*: The characteristic should be sufficiently invariant over a period of time.
- *Collectability*: The characteristic can be measured quantitatively.

However, in a practical biometric system (i.e., a system that employs biometrics for personal recognition), there are a number of other issues that should be considered including:

- *Performance*. Which refers to the achievable recognition accuracy and speed, the resources required to achieve the desired recognition accuracy and speed, as well as the operational and environmental factors that affect the accuracy and speed?
- *Acceptability*. Which indicates the extent to which people are willing to accept the use of a particular biometric identifier (characteristic) in their daily lives?
- *Circumvention*. Which reflects how easily the system can be fooled using fraudulent methods?

1.8 Biometrics Operating Modes

A biometric system is essentially a pattern recognition system that may operate either in verification mode or identification mode:

1.8.1 Verification Mode

In this mode, the system validates a person's identity by comparing the captured biometric data with her own biometric template(s) stored system database. In such a system, an individual who desires to be recognized claims an identity, usually via a PIN (Personal Identification Number), a user name, a smart card, etc. and the system conducts a one-to-one comparison to determine whether the claim is true or not. Identity verification is typically used for positive recognition, where the aim is to prevent multiple people from using the same identity [9].

1.8.2 Identification Mode

In identification mode, the system recognizes an individual by searching the templates of all the users in the database for a match. Therefore, the system conducts a one-to-many comparison to establish an individual's identity (or fails if the subject is not enrolled in the system database) without the subject having to claim an identity. Identification is a critical component in negative recognition applications where the system establishes whether the person is who she (implicitly or explicitly) denies to be. The purpose of negative recognition is to prevent a single person from using multiple identities. Identification may also be used in positive recognition for convenience (the user is not required to claim an identity). While traditional methods of personal recognition such as passwords, PINs, keys, and tokens may work for positive recognition, negative recognition can only be established through biometrics.

1.9 Applications of Biometrics

The areas of biometrics applications are dynamically extended. It is possible to explain partially that problems of information safety became priority for the present stage of information technology development. The hopes on biometrics are in many respects justified and their use frequently gives good results. The modern biometrics applications mainly concentrate in the areas of medical, law and order, banking and finance, immigration control, visual and voice communications, access control as indicated in Figure 1.3.

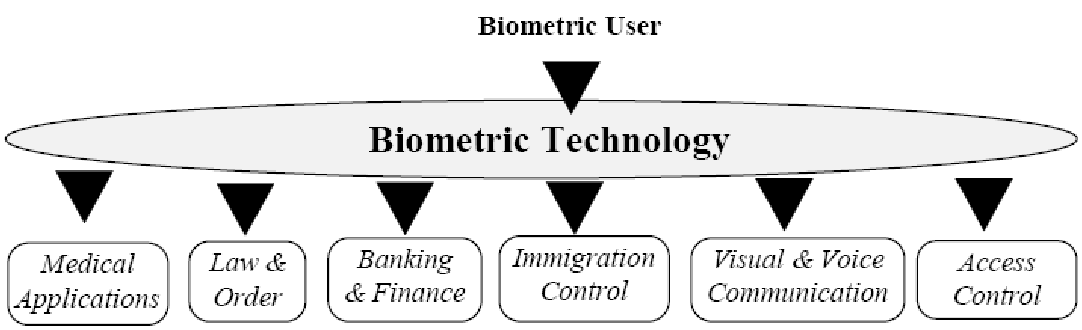


Figure 1.3. Biometrics Main Applications

Traditionally, commercial applications have used knowledge-based systems (e.g., PINs and passwords), government applications have used token-based systems (e.g., ID cards and badges), and forensic applications have relied on human experts to match biometric features. Biometric systems are being increasingly deployed in large scale civilian applications. At the some airport, iris scanning is carried out to speed up the passport and visa control procedures [10]. Passengers enrolled in this scheme insert their card at the gate and look into a camera; the camera acquires the image of the traveler's eye and processes it to locate the iris, and compute the iris code. The computed iris code is compared with the data residing in the card to complete user verification. A similar scheme is also being used to verify the identity of airport employees working in high-security areas. Thus, biometric systems can be used to enhance user convenience while improving security.

1.10 Comparison of Various Biometrics

A number of biometric characteristics exist and are in use in various applications. Each biometric has its strengths and weaknesses, and the choice depends on the application. No single biometric is expected to effectively meet the requirements of all the applications. In other words, no biometric is "optimal". The match between a specific biometric and an application is determined depending upon the operational mode of the application and the properties of the biometric characteristic. A brief introduction of the commonly used biometrics is given below:

1.10.1 Hand Writing and Signature Recognition

Now-a-days the automatic investigation of handwritten objects has been widely used to confirm the document authenticity in the financial sphere, to solve the expert problems in criminology, to diagnose the physical and psychic state of patients in medicine, to make the psychological individual analysis in psychology and others. Computer analysis of handwriting and signature is very interesting research area. Most of the work done in this area has made use of real-time input (i.e., handwriting objects are input to the computer by means of light pen, table, and so on), so that the time sequence of strokes in the signature and handwriting is known. Automatic reading of unknown signature and handwriting is a difficult problem. Signature and handwriting varies greatly, even for a single individual. The task of handwriting objects verification is to determine whether a given signature as shown in Figure 1.4 or handwriting is genuine or forgery. The forged handwriting object has the proper shape, but differs from genuine object in the quality of the strokes. Free-hand forgeries, on the other hand, differ from a genuine signature with respect to the values of various size and shape features. The modeling of handwriting objects is a highly nontrivial task, since even genuine signatures and handwriting can be quite variable in size and shape, and template matching is not a viable approach to verification. Certain features of these objects are, however, believed to be relatively invariant for a given writer. These include ratios of small letter height to signature width, of tall letter height to small letter height, distances between letters, and so on. The signatures are a behavioral biometric that change over a period of time and are influenced by physical and emotional conditions of the signatories.

1.10.2 Keystroke Dynamics

Everybody has his own typing rhythm and it is this rhythm that keystroke dynamics utilizes. The process of typing the PIN can be broken down into quantifiable components, such as latency time, key press force, key press duration and key press displacement which can be evaluated and used to verify the identity of a person. As the PIN signature is like the written signature that differs slightly with every execution, a neural-fuzzy application is devised to verify the PIN signature input against the reference profile. Like signature dynamics, keystroke dynamics are difficult to reproduce.

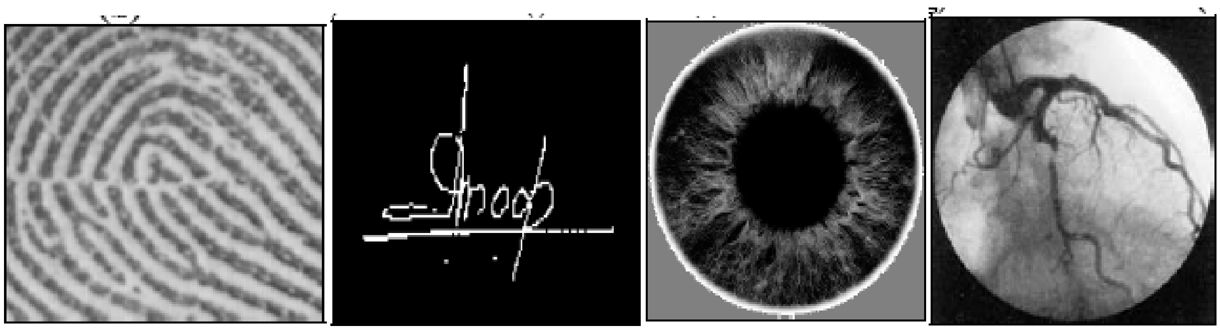


Figure 1.4. Few Examples of Biometrics (Finger print, Hand writing, Iris and Blood vessels in Retina)

1.10.3 Fingerprints Recognition

Humans have used fingerprints for personal identification for many centuries and the matching accuracy using fingerprints has been shown to be very high [11]. A fingerprint as shown in Figure 1.4 is the pattern of ridges and valleys on the surface of a fingertip, the formation of which is determined during the first seven months of fetal development. Fingerprints of identical twins are different and so are the prints on each finger of the same person. The accuracy of the currently available fingerprint recognition systems is adequate for verification systems and small- to-medium scale identification systems involving a few hundred users. Multiple fingerprints of a person provide additional information to allow for large-scale recognition involving millions of identities. One problem with the current fingerprint recognition systems is that they require a large amount of computational resources, especially when operating in the identification mode.

1.10.4 Recognition Through Eye

Traditionally two ways are used to identify the eyes. These ways differ in accordance with the zones of eye, from which the attributes are obtained: (i) the blood vessels on the retina as shown in Figure 1.4 and (ii) iris as pattern is shown in Figure 1.4. Iris is the annular region of the eye bounded by the pupil and the sclera (white of the eye) on either side. The visual texture of the iris is formed during fetal development and stabilizes during the first two years of life. The complex iris texture carries very distinctive information useful for personal recognition. The accuracy and speed of currently deployed iris-based recognition systems is promising and point to the feasibility of large-scale identification systems based on iris information. Each iris is distinctive and, like fingerprints, even the irises of identical twins are different. It is extremely difficult to surgically tamper the texture of the iris.

Retinal vasculature is rich in structure and is supposed to be a characteristic of each individual and each eye. It is claimed to be the most secure biometric, since it is not easy to change or replicate the retinal vasculature. The image acquisition requires a person to peep into an eye-piece and focus on a specific spot in the visual field, so that a predetermined part of the retinal vasculature could be imaged. The image acquisition involves cooperation of the subject, entails contact with the eyepiece, and requires a conscious effort on the part of the user. All these factors adversely affect the public acceptability of retinal biometric.

1.10.5 Hand and Finger Geometry

Hand geometry recognition systems are based on a number of measurements taken from the human hand, including its shape, size of palm, and lengths and widths of the fingers. The geometry of the hand is not known to be very distinctive and hand geometry-based recognition systems cannot be scaled up for systems requiring identification of an individual from a large population. Further, hand geometry information may not be invariant during the growth period of children. In addition, an individual's jewelry (e.g., rings) or limitations in dexterity (e.g., from arthritis), may pose further challenges in extracting the correct hand geometry information. The physical size of a hand geometry-based system is large, and it cannot be embedded in certain devices like laptops. There are verification systems available that are based on measurements of only a few fingers (typically, index and middle) instead of the entire hand. These devices are smaller than those used for hand geometry, but still much larger than those used in some other biometrics (e.g., fingerprint, face, voice).

1.10.6 Voice Recognition

Voice is a combination of physiological and behavioral biometrics. The features of an individual's voice are based on the shape and size of the appendages (e.g., vocal tracts, mouth, nasal cavities, and lips) that are used in the synthesis of the sound. These physiological characteristics of human speech are invariant for an individual, but the behavioral part of the speech of a person changes over time due to age, medical conditions (such as common cold), emotional state, etc. Voice is also not very distinctive and may not be appropriate for large-scale identification. A text-dependent voice recognition system is based on the utterance of a fixed predetermined phrase. A text-independent voice recognition system recognizes the speaker independent of what he/she speaks.

1.10.7 Gait

Gait is the peculiar way one walks and is a complex spatio-temporal biometric as shown in Figure 1.5. Gait is not supposed to be very distinctive, but is sufficiently discriminatory to allow verification in some low-security applications. Gait is a behavioral biometric and may not remain invariant, especially over a long period of time, due to fluctuations in body weight, major injuries involving joints or brain, or due to inebriety. Acquisition of gait is similar to acquiring a facial picture and, hence, may be an acceptable biometric. Since gait-based systems use the video-sequence footage of a walking person to measure several different movements of each articulate joint, it is input intensive and computationally expensive.

1.10.8 Facial and Hand Vein Infrared

The pattern of heat radiated by human body is a characteristic of an individual and can be captured by an infrared camera in an unobtrusive way much like a regular (visible spectrum) photograph as shown in Figure 1.5 The technology could be used for covert recognition. A thermogram-based system does not require contact and is non-invasive, but image acquisition is challenging in uncontrolled environments, where heat emanating surfaces (e.g., room heaters and vehicle exhaust pipes) are present in the vicinity of the body. A related technology using near infrared imaging is used to scan the back of a clenched fist to determine hand vein structure. Infrared sensors are prohibitively expensive which is a factor inhibiting wide spread use of the thermograms.

1.10.9 Ear Recognition

It has been suggested that the shape of the ear and the structure of the cartilaginous tissue of the pinna are distinctive as shown in Figure 1.5. The ear recognition approaches are based on matching the distance of salient points on the pinna from a landmark location on the ear. The features of an ear are not expected to be very distinctive in establishing the identity of an individual.

1.10.10 DNA

Deoxyribo Nucleic Acid (DNA) is the one-dimensional ultimate unique code for one’s individuality, except for the fact that identical twins have identical DNA patterns. It is however, currently used mostly in the context of forensic applications for person recognition.

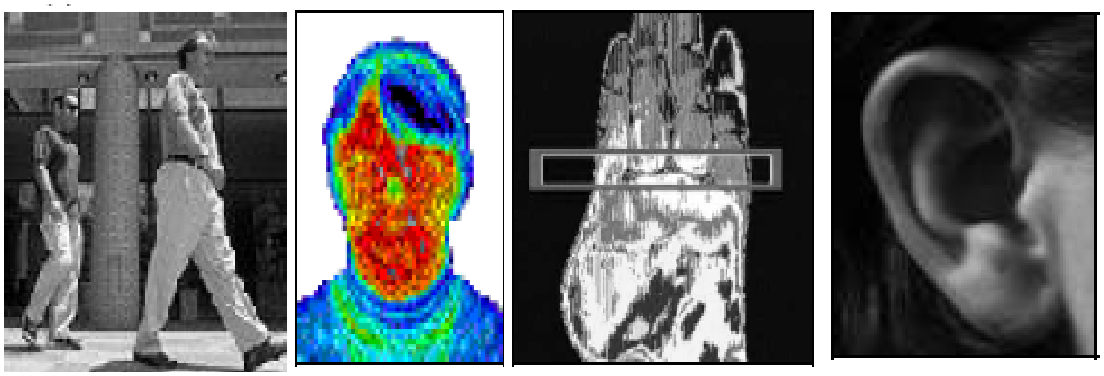


Figure 1.5. Biometrics Examples (Gait, Facial, hand vein infrared and ear geometry)

Three issues limit the utility of this biometrics for other applications: (i) contamination and sensitivity: it is easy to steal a piece of DNA from an unsuspecting subject that can be subsequently abused for an ulterior purpose; (ii) automatic real-time recognition issues: the present technology for DNA matching requires cumbersome chemical methods (wet processes) involving an expert's skills and is not geared for on-line non-invasive recognition and (iii) privacy issues: information about susceptibilities of a person to certain diseases could be gained from the DNA pattern and there is a concern that the unintended abuse of genetic code information may result in discrimination.

1.10.11 Face Recognition

Researchers in computer vision and pattern recognition have worked on automatic techniques for recognizing human faces for the last 20 years. Face recognition is a non-intrusive method, and facial images are probably the most common biometric characteristic used by humans to make a personal recognition. The applications of facial recognition range from a static, controlled "mug-shot" verification to a dynamic, uncontrolled face identification in a cluttered background (e.g., airport). The most popular approaches to face recognition 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 verification performance of the face recognition systems that are commercially available is reasonable [12], they impose a number of restrictions on how the facial images are obtained, sometimes requiring a fixed and simple background or special illumination. These systems also have difficulty in recognizing a face from images captured from two drastically different views and under different illumination conditions. 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 [13]. In order to ensure that a facial recognition system works 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; and (iii) recognize the face from a general viewpoint (i.e., from any pose). In facial recognition system, computers perform three classical tasks: verification (the system attempts to match a live face with a specific reference digital image), recognition (the system try to match a live face with any saved faces in a central computer database), and locating the face within the image. One of the problem in face recognition is face images have varying pose and tilt. For finding its solution, we use: (i) representation of faces with templates from multiple model views that cover different poses from the viewing sphere, (ii) recognition of a novel view, the recognizer locates the eyes and nose features, uses these locations to geometrically register the input model views, (iii) using the correlation on model templates to find the best match in the database of people. One of the very important steps of this process is to determine the best facial features to discriminate the features of one face from those of another. An edge enhancing preprocessor and some separately trained back propagation Neural Networks (NN) are often applied to address face recognition problem. The applicability of a specific biometric technique depends heavily on the requirements of the application domain. No single technique can out-perform all the others in all operational environments. In this sense, each biometric technique is admissible and there is no optimal biometric characteristic. For example, it is well known that face, fingerprint and iris based techniques are more accurate than the voice-based technique.

Figure 1.6. Face Images of FERET Database with Varying Poses

1.11 Limitations of (Unimodal) Biometric Systems

The successful installation of biometric systems in various civilian applications does not imply that biometrics is a fully solved problem. It is clear that there is plenty of scope for improvement in biometrics. Researchers are not only addressing issues related to reducing error rates, but they are also looking at ways to enhance the usability of biometric systems. Biometric systems that operate using any single biometric characteristic have the following limitations:

1.11.1 Noise in Sensed Data

The sensed data might be noisy or distorted. A fingerprint with a scar, or a voice altered by cold are examples of noisy data. Noisy data could also be the result of defective or improperly maintained sensors (e.g., accumulation of dirt on a fingerprint sensor) or unfavorable ambient conditions (e.g., poor illumination of a user's face in a face recognition system). Noisy biometric data may be incorrectly matched with templates in the database resulting in a user being incorrectly rejected.

1.11.2 Intra-Class Variations

The biometric data acquired from an individual during authentication may be very different from the data that was used to generate the template during enrollment, thereby affecting the matching process. This variation is typically caused by a user who is incorrectly interacting with the sensor or when sensor characteristics are modified (e.g., by changing sensors - the sensor interoperability problem) during the verification phase. As another example, the varying psychological makeup of an individual might result in vastly different behavioral traits at various time instances.

1.11.3 Distinctiveness

While a biometric trait is expected to vary significantly across individuals, there may be large inter-class similarities in the feature sets used to represent these traits. This limitation restricts the discriminability provided by the biometric trait. Golfarelli et al. [14] have shown that the information content (number of distinguishable patterns) in two of the most commonly used representations of hand geometry and face are only of the order of 10^5 and 10^3, respectively. Thus, every biometric trait has some theoretical upper bound in terms of its discrimination capability.

1.11.4 Non-Universality

While every user is expected to possess the biometric trait being acquired, in reality it is possible for a subset of the users to not possess a particular biometric. A fingerprint biometric system, for example, may be unable to extract features from the fingerprints of certain individuals, due to the poor quality of the ridges. Thus, there is a Failure To Enroll (FTE) rate associated with using a single biometric trait. It has been empirically estimated that as much as 4% of the population may have poor quality fingerprint ridges that are difficult to image with the currently available fingerprint sensors and result in FTE errors.

1.11.5 Spoof Attacks

An impostor may attempt to spoof the biometric trait of a legitimate enrolled user in order to circumvent the system. This type of attack is especially relevant when behavioral traits such as signature [15] and voice [16] are used. However, physical traits are also susceptible to spoof attacks. For example, it has been demonstrated that it is possible (although difficult and cumbersome and requires the help of a legitimate user) to construct artificial fingers/fingerprints in a reasonable amount of time to circumvent a fingerprint verification system [17].

1.12 Overview of Thesis

The thesis is organized in seven chapters. Chapter two focuses on fundamentals of face recognition, its most popular techniques and various applications.

Chapter three discusses various newly developed preprocessing techniques for face images. The preprocessing is necessary to prepare the face images for subsequent face recognition operations. The face databases employed for the experiments are also discussed.

Chapter four focuses on dimension reduction of feature vectors through reduction of image resolution. A reduced dimension feature vector facilitates computationally efficient face recognition operations. The reduction in image resolution is achieved by low pass filtering and subsequent decimation operations. Spatial domain template matching is then used for face recognition.

Chapter five deals with transform domain face recognition techniques for reduced resolution images. The various transforms used are Fourier Transform, Discrete Cosine Transform and Wavelet Transform. The different feature extraction methods in transform domain are employed and their effect on face recognition performance is investigated. Moreover, effect of varying image resolution on recognition in transform domain is examined.

Chapter six extends the discussion of image resolution reduction from square to hexagonal image framework. Based on diagonal grow and butterfly structure methodology, a new technique for sampling and indexing of hexagonal structure is proposed.

Concluding remarks and some future recommendations are given in Chapter seven.

CHAPTER 2

FACE RECOGNITION FUNDAMENTALS AND CHALLENGES

2.1 Introduction

Biometrics are automated methods of recognizing an individual based on their physical or behavioral characteristics. There are many different types available commercially or being researched for future use including Body Odor, Ear Biometrics, Face Recognition, Facial Thermography, Fingerprint, Gait, Hand Geometry, Iris Recognition, Keystroke Dynamics, Signature Dynamics, Speaker Recognition and Vein [18].

Facial recognition is by no means a perfect technology and much technical work has to be done before it becomes a truly viable tool to counter terrorism and crime. But the technology is getting better and there is no denying its tremendous potential. In the meantime, society has time to decide how it wants to use this new technology. By implementing reasonable safeguards, society can harness the power of the technology to maximize its public safety benefits while minimizing the intrusion on individual privacy.

Face recognition uses the visible physical structure of an individual's face for recognition purposes. In order to determine if a biometric works and which system(s) should be deployed, one needs to first properly define the application of interest. This definition needs to be as specific as possible because even a small variance can sometimes significantly alter anticipated performance. Questions to ask when defining a potential face recognition application include:

- Identification, verification or watch list mode of operation
- The size of the database for identification or watch list
- Demographics of the anticipated users (age, sex, etc.)
- Lighting conditions – indoor/outdoor
- System to be installed overtly or covertly
- Anticipated user behavior

- Life of the user/image enrolled
- Required throughput rate
- Exception handling cases to be handled for a given period of time
- Most vital parameter (identification: rank or identification rate; verification: false alarm or probability of verification; watch list: false alarm or correct alarm)
- Minimum accuracy requirements

2.2 Need for Face Recognition

Although the concept of recognizing someone from facial features is intuitive, facial recognition, as a biometric, makes human recognition a more automated, computerized process. What sets apart facial recognition from other biometrics is that it can be used for surveillance purposes. For example, public safety authorities want to locate certain individuals such as wanted criminals, suspected terrorists, and missing children. Facial recognition may have the potential to help the authorities with this mission. Facial recognition offers several advantages. The system captures faces of people in public areas, which minimizes legal concerns. Moreover, since faces can be captured from some distance away, facial recognition can be done without any physical contact. This feature also gives facial recognition a clandestine or covert capability. For any biometric system to operate, it must have records in its database against which it can search for matches. Facial recognition is able to leverage existing databases in many cases. For example, there are high quality mugshots of criminals readily available to law enforcement. Similarly, facial recognition is often able to leverage existing surveillance systems such as surveillance cameras or Closed Circuit Television (CCTV).

2.3 Five Steps to Facial Recognition

As a biometric, facial recognition is a form of computer vision that uses faces to attempt to identify a person or verify a person’s claimed identity. Regardless of specific method used, facial recognition is accomplished in a five step process [19] as shown in Figure 2.1.

2.3.1 Capture Image

First, an image of the face is acquired. This acquisition can be accomplished by digitally scanning an existing photograph or by using an electro-optical camera to acquire a live picture of a subject. As video is a rapid sequence of individual still images which can also be used as a source of facial images.

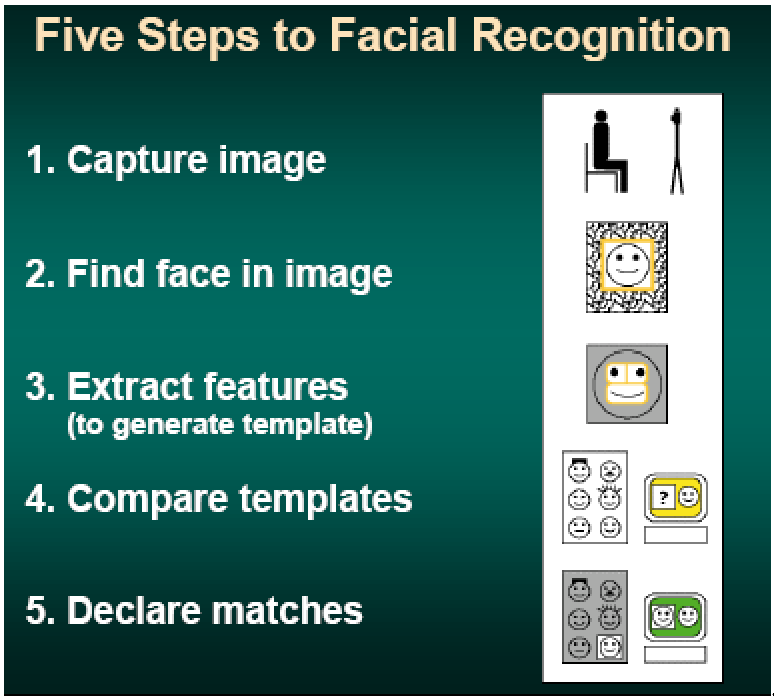


Figure 2.1. Steps of Facial Recognition

2.3.2 Find Face in Image

Software is employed to detect the location of any faces in the acquired image. This task is difficult, and often generalized patterns of what a face "looks like" (two eyes and a mouth set in an oval shape) are employed to pick out the faces.

2.3.3 Extract Features

Facial recognition analyzes the spatial geometry of distinguishing features of the face. Different vendors use different methods to extract the identifying features of a face.

Thus, specific details on the methods are proprietary. The most popular method is called Principal Components Analysis (PCA), which is commonly referred to as the eigenface method. PCA has also been combined with neural networks and local feature analysis in efforts to enhance its performance. Template generation is the result of the feature extraction process. A template is a reduced set of data that represents the unique features of an enrollee's face. It is important to note that because the systems use spatial geometry of distinguishing facial features, they do not use hairstyle, facial hair, or other similar factors.

2.3.4 Compare Templates

This step compares the generated template with those in a database of known faces. In an identification application, this process yields scores that indicate how closely the generated template matches each of those in the database. In a verification application, the generated template is only compared with one template in the database – that of the claimed identity.

2.3.5 Declare Matches

The final step determines whether scores produced in 4th step are high enough to declare a match. The rules governing the declaration of a match are often configurable by the end user, so that he or she can determine how the facial recognition system should behave based on security and operational considerations.

2.4 Human and Technical Difficulties with Facial Recognition

People are generally very good at recognizing faces that they know. However, people experience difficulties when they perform facial recognition in surveillance or watch post scenario. Combined with relatively short attention spans, it is difficult for humans to pick out unfamiliar faces. In addition to unfamiliar face recognition problems, the ability of human beings to detect critical signals drops rapidly from the start of a task and stabilizes at a significantly lower level within 25 to 35 minutes. Thus the ability of people to focus their attention drops significantly after only half an hour.

Machines also experience difficulties when they perform facial recognition in surveillance or watch post scenario. Dr. Wayman [20] has explained that performing facial recognition processes with relatively high fidelity and at long distances remains technically challenging for automated systems. At the most basic level, detecting whether a face is present in a given electronic photograph is a difficult technical problem. Dr. Wayman [21] has noted that subjects should ideally be photographed under tightly controlled conditions. For example, each subject should look directly into the camera and fill the area of the photo for an automated system to reliably identify the individual or even detect his/her face in the photograph. Thus, while the technology for facial recognition systems shows promise, it is not yet considered fully mature.

2.5 Typical Face Recognition System

A generalized face recognition model is shown in Figure 2.2.

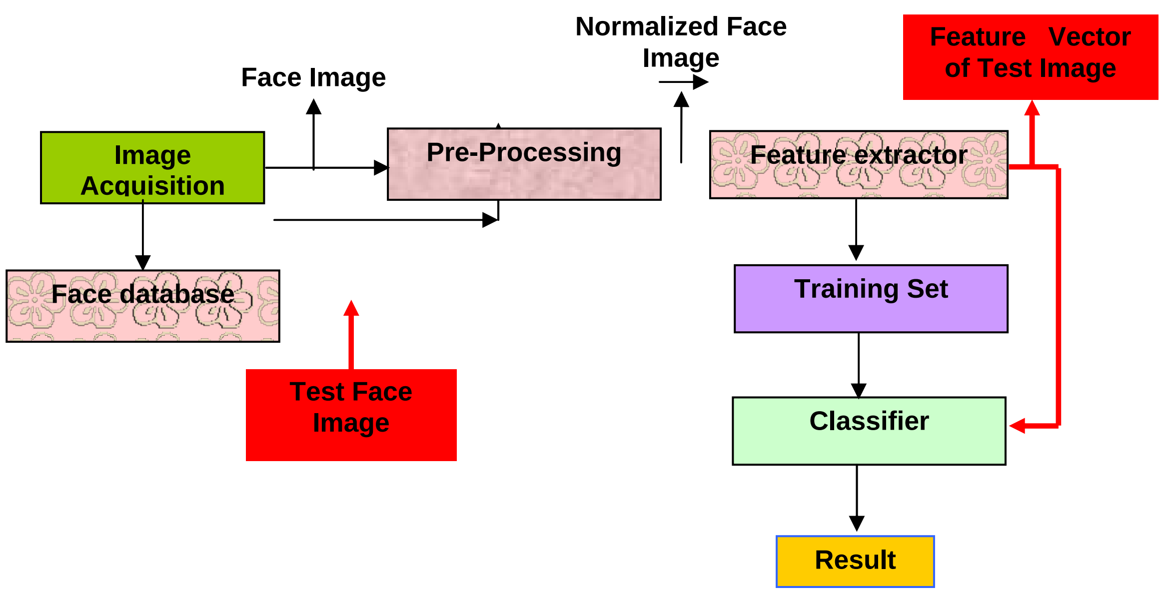


Figure.2 2. Typical Face Recognition System

There are six main functional blocks whose functions are described below:

2.5.1 The Acquisition Module

This is the entry point of the face recognition process. It is the module where the face image under consideration is presented to the system. In other words, the user is asked to present a face image to the face recognition system in this module. An acquisition module can request 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.

2.5.2 The Pre-processing Module

In this module, by means of early vision techniques, face images are normalized and if desired, they are enhanced to improve the recognition performance of the system. Some or all of the following pre-processing steps may be implemented in a face recognition system.

2.5.2.1 Image size normalization. It is usually done to change the acquired image size to a default image size such as 128 x 128, on which the face recognition system operates. This is mostly encountered in systems where face images are treated as a whole.

2.5.2.2 Histogram equalization. It is usually done on too dark or too bright images in order to enhance image quality and to improve face recognition performance. It modifies the dynamic range (contrast range) of the image and as a result, some important facial features become more apparent.

2.5.2.3 Median filtering. For noisy images especially obtained from a camera or from a frame grabber, median filtering can clean the image without loosing information.

2.5.2.4 High-pass filtering. Feature extractors that are based on facial outlines, may benefit the results that are obtained from an edge detection scheme. High-pass filtering emphasizes the details of an image such as contours which can dramatically improve edge detection performance.

2.5.2.5 Background removal. In order to deal primarily with facial information itself, face background can be removed. This is especially important for face recognition systems where entire information contained in the image is used. It is obvious that, for background removal, the preprocessing module should be capable of determining the face outline.

2.5.2.6 Translational and rotational normalizations. In some cases, it is possible to work on a face image in which the head is somehow shifted or rotated. The head plays the key role in the determination of facial features. Especially for face recognition systems that are based on the frontal views of faces, it may be desirable that the pre processing module determines and, if possible normalizes the shifts and rotations in the head position.

2.5.2.7 Illumination normalization. Face images taken under different illuminations can degrade recognition performance, especially for face recognition systems in which entire face information is used for recognition.

2.5.3 The Feature Extraction Module

After performing some pre-processing (if necessary), the normalized face image is presented to the feature extraction module in order to find the key features that are going to be used for classification. In other words, this module is responsible for composing a feature vector that is well enough to represent the face image.

2.5.4 The 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.

2.5.5 Training Set

Training sets are used during the "learning phase" of the face recognition process. The feature extraction and the classification modules adjust their parameters to achieve optimum recognition performance by making use of training sets.

2.5.6 Face Library or Face Database

After being classified as "unknown", face images can be added to a library (or to a database) with their feature vectors for later comparisons. The classification module makes direct use of the face library.

2.6 Brief Overview of Face Recognition Approaches

Face recognition approaches can be grouped in to three categories which include Holistic [21,22, 23, 24], Feature Based [25,26,27] and Hybrid [28] approaches. Holistic face recognition utilizes global information from faces to perform face recognition. The global information from faces is fundamentally represented by a small number of features which are directly derived from the pixel information of face images. These small numbers of features distinctly capture the variance among different individual faces and therefore are used to uniquely identify individuals. Feature based face recognition uses a priori information or local features of faces to select a number of features to uniquely identify individuals. Local features include the eyes, nose, mouth, chin and head outline, which are selected from face images. The general approaches to feature based face recognition are concerned with using a priori information of the face to find local face features [29]. This implementation consisted of two parts: the first part was designed to identify individual features, such as the eyes (general location of the eyes), eye (iris and surrounding white around the iris), chin, cheek, hair, jaw-line, mouth, mouth bits (edges and outline of the lips), head outline and the nose; the second part refined the features found from the first part by using a priori information to locate 40 pre-defined face features. Specifically, the process of finding these 40 pre-defined features were to initially find the head outline within face images. By finding the head outline, this facilitated finding other face features.

Hybrid approaches have a special status among face recognition systems as they combine different recognition approaches in an either serial or parallel order to overcome the shortcomings of the individual components. Different recognition approaches succeed and fail at widely different viewing and illumination conditions, which usually cannot be accommodated by day-to-day recognition systems. Due to this dilemma, it seems obvious to run various individual recognition classifiers on a problem leading to an individual ranking of the results of every process, and to design a classification scheme to assess an overall recognition result. In contrast to this parallel approach of equivalent face classifiers, a serial one is also conceivable where the output of one classifier is input to the next one. The latter approach can even go along with hybrid learning if the classifiers require training.

In preceding section, an overview of few main approaches of all three categories is presented.

2.6.1 Nearest - Neighbor Template Matching

Nearest-Neighbor template matching is the most native method of face recognition [30]. Individuals are represented by a set of complete images of their face. An input face is compared with each image in the database. The image for which the template matching gives the highest response is considered to be the match, given that it is significant enough to represent a match. Though simple this method has some shortcomings. Template matching has an inability to deal with changes in intrinsic and extrinsic factors. A large database, images of all possible factors, would be required to handle such variations. Not only would this be time consuming to initially form the database but would increase the computational intensity of the detection phase resulting in an extremely slow system. With an image of size 92x112 there are 10,304 pixels. Assume comparing one image with 10,304 pixels within a database of 40 people with 5 images each. This results in a comparison of 10,304 x 40 x 5 = 2,060,800 pixels. This would also require a large amount of memory space and would result in an inefficient system.

2.6.2 Principal Component Analysis

The Principal Component Analysis (PCA) or Hotelling transform, is a well known method for dimension reduction. PCA has been used widely in computer visions applications [31,32]. In terms of information theory, the purpose of using PCA is to extract the relevant information in a face image, encode it as efficiently as possible, and compare one face encoding with a database of similarly encoded models. A simple approach to extract the information contained in an image of face is to somehow capture the variation in a collection of face images independent of judgment of features and use this information to encode and compare individual faces.

In mathematical terms, the purpose of using PCA is to find the principal components of distribution of faces, or the eigenvectors of the covariance matrix of the set of face images, treating each image as a point (vector) in a very high dimensional space. The eigenvectors are ordered, each one accounting for a different amount of 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. Each eigenvector appears as a ghostly face. 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.

PCA reduces the dimension by creating a new co-ordinate system for the given data set. It attempts to produce axes, which have a large percentage of variance along them, such that it is possible to ignore dimensions/axes without losing much information. In general, a reduction in the dimensionality of the input space will be accompanied by a loss of some of the information which discriminated between different classes. The goal in dimensionality reduction is therefore to preserve as much of the relevant information as possible. The procedure revolves around, techniques for combining inputs together to make a smaller set of features. The procedures rely entirely on the input data itself without reference to the corresponding target or test data.

2.6.3 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is very similar to PCA. They are both linear transforms. However, LDA attempts to maximize the difference between images [33, 34].

LDA makes use of PCA as a pre-calculation phase. The dimension reduction is achieved by a combination of PCA and LDA. There are two types of scatter matrices sought in LDA: Within Class Scatter Matrix and Between Class Scatter Matrix. LDA can only separate classes that are linearly separated. If the faces were not linearly grouped i.e. the classes / subjects can not be separated with a linear projection, then LDA will not perform well.

2.6.4 Independent Component Analysis

PCA can be derived as a special case of Independent Component Analysis (ICA) [35]. While PCA de-correlates the input data using second-order statistics and thereby generates compressed data with minimum mean-squared re-projection error, ICA minimizes both second-order and higher-order dependencies in the input. It is intimately related to Blind Source Separation (BSS) problem, where the goal is to decompose an observed signal into linear combination of unknown independent signals [36].

2.6.5 Discrete Cosine Transformation

An alternative approach for face recognition tasks is to use a Discrete Cosine Transformation (DCT), it is another way of information packing. It is similar in theory to PCA but with a major difference-PCA is data dependent and is a nontrivial computational task whereas 2D discrete cosine transformations can be calculated fairly quickly. This makes DCT an extremely competitive technique in terms of computational complexity. The cosine transform is similar to the Fourier transform. They both use sinusoidal basis functions. Cosine transform basis functions are non-complex; they only use cosine functions, and not sine functions. The definition of a discrete cosine transform is:

$$C(u,v) = \frac{2}{\sqrt{MN}} \alpha(u)\alpha(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \cos\left[\frac{(2x+1)u\pi}{2M}\right] \cos\left[\frac{(2y+1)v\pi}{2N}\right] \tag{2.1}$$

where f(x,y) is Input image

C(u,v) is DCT Transformed spectrum of Image

Inverse transform

$$f(x,y)=\frac{2}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\alpha(u)\alpha(v)C(u,v)\cos\left[\frac{(2x+1)u\pi}{2M}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right] \quad (2.2)$$

$$\alpha(w)=\begin{cases}\frac{1}{\sqrt{2}}\ ,if\ \ w=0\\ 1\ ,\ otherwise\end{cases}$$

2.6.6 Non-Negative Matrix Factorization

Non-negative matrix factorization is differentiated from other methods by use of its non-negativity constraints. These constraints lead to a parts-based representation because they allow only additive, not subtractive combinations. When non-negative matrix factorization is implemented as a neural network, parts based representation emerges by virtue of two properties: the firing rates of neurons are never negative and synaptic strengths do not change sign [37].

2.6.7 Neural Networks

From the point of view of face recognition, a set of optimal boundaries between different classes should be estimated by neural networks. Conversely, from the point of view of neural networks, the neural networks are regarded as a mapping from the feature hyperspace to the classes [38]. Each pattern is represented by a real vector and each class is assigned for a suitable code. Therefore, the number of inputs is set equal to that of features (i.e. the dimension of the input space) and the number of outputs is set equal to that of classes. It is difficult to select the hidden nodes. Different approaches revolve around increasing or decreasing the complexity of the architecture.

2.6.8 Support Vector Machines

The use of Support Vector Machines (SVM) [39] has previously been used for general pattern recognition, but also applied to face recognition by Guo [40] and had used SVM in conjunction with a binary decision tree, which they found comparable improvements to the Eigenface approach. Also, Xi et al [41] reported using SVM for face recognition where they were able to accurately and autonomously extract features from faces. Tefas et al. [42] had used SVM with the elastic bunch graph matching approach and reported achieving a lower error rate for recognizing faces when compared to the standard elastic bunch graph matching approach.

2.6.9 Fourier Transform

Fourier Transform (FT) is another approach applied to face recognition where recognition is done by finding the closest match between feature vectors containing the Fourier coefficients at selected frequencies. An efficient implementation of Fourier Transform is Fast Fourier Transform (FFT) which is used practically. Transform is defined as:

Colmenarez and Huang [43] proposed using the FFT approach to represent the gallery set as a number of templates. Face recognition for this approach was achieved by minimizing the Euclidean distance amongst the peaks found between comparing a gallery face to a probe face. Fourier Transform is defined as:

$$F(u) = \frac{1}{M}\sum_{x=0}^{M-1} f(x)e^{-j2\pi ux/M} \tag{2.3}$$

for u=0,1,2,.....M-1

$$f(x) = \frac{1}{M}\sum_{u=0}^{M-1} F(u)e^{j2\pi ux/M} \tag{2.4}$$

for x=0,1,2,.....M-1

2.6.10 Elastic Bunch Graph Matching

Elastic Bunch Graph Matching (EBGM) recognizes faces by matching the probe set represented as the input face graphs, to the gallery set that is represented as the model face graph. Fundamental to the EBGM is the concept of nodes. Essentially, each node of the input face graph is represented by a specific feature point of the face. For example, a node represents an eye and another node represents the nose and the concept continues for representing the other face features. Therefore, the nodes for the input face graph are interconnected to form a graph like data structure which is fitted to the shape of the face as illustrated in Figure 2.3.

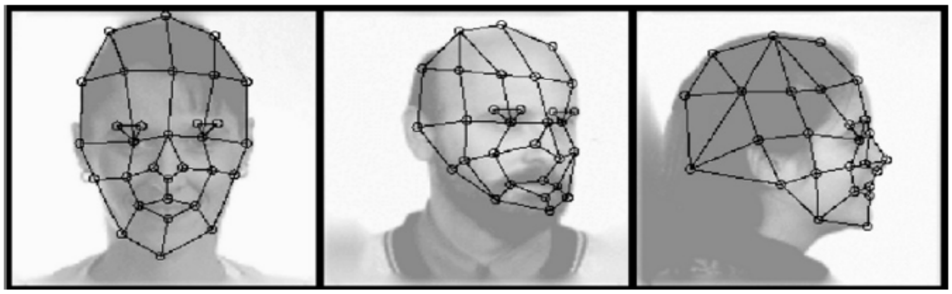


Figure 2.3.Elastic Bunch Graph of Facial Features

In contrast, the model face graph represents the gallery set only used one model face graph to represent the entire gallery set. However, the model face graph can be conceptually thought of as a number of input face graphs stacked on top of each other and concatenated to form one model face graph, with the exception that this is applied to the gallery set instead of the probe set. Therefore, this would allow the grouping of the same types of face features from different individuals. For example, the eyes of different individuals could be grouped together to form the eye feature point for the model face graph and the noses of different individuals can be grouped together to form the nose feature point for the model face graph. An illustration of the model face graph is shown in Figure 2.4. Given the definition for the input face graph and model face graph, to determine the identity for the input face graph is to achieve the smallest distance in relation to the model face graph for a particular gallery face. The distance is determined by the node similarity measure for the input face graphs to the model face graph.

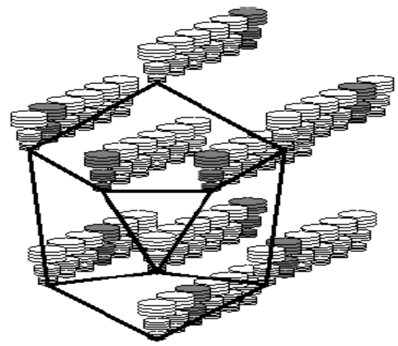


Figure 2.4. Illustration of Model Face Graph

2.6.11 Face Recognition Using Wavelets

Wavelet transform techniques are not too old and these techniques are being used in modern signal and image processing including multiresolution analysis, sound synthesis, computer vision, graphics and image compression [44]. Wavelet transform techniques achieve optimal decomposition without affecting much the image quality. At the same time wavelet transform and wavelet packet analysis have provided a new subspace for face recognition. Fotlyniewiez [45] proposed an automatic face recognition using nonlinear filtering to enhance intrinsic features of face and used a high order neural network classifier for training and recognition of faces. Lee and Chung [46] employed the wavelet-based Fisher Linear Discriminant (FLD) recognition process. Zhu and Orchad [47] captured local discriminative features in the space frequency domain for face detection using wavelet packet analysis. Ma and Tang [48] used discrete wavelet face graph matching approach for the purpose. Liu [49] used Haar wavelet for effective human face detection. Yang [50] illustrates an application of nonlinear wavelet approximation to recognize faces and the advantages of nonlinear wavelet approximation are compared with its linear counterpart. Wiskott [51] used labeled graph based on Gabor wavelet transform for face recognition application.

2.7 Face Recognition Challenges

Face recognition involves computer recognition of personal identity based on geometric or statistical features derived from face images [52, 53, 54]. Even though humans can detect and identify faces in a scene with little or no effort, building an automated system that accomplishes such objectives is, however, very challenging. The challenges are even more profound when one considers the large variations in the visual stimulus due to illumination conditions, viewing directions or poses, facial expression, aging and most importantly varying image resolution.

A good face recognition system should exhibit the following features:

- It should not assume any restriction on pose and illumination conditions (i.e., the gallery and probe images can be taken under different poses and different illumination).
- It should be able to extract best features required for recognition and provide expression and facial tilt compensation.
- It should be able to achieve best image resolution against best recognition results.

2.8 Image Illumination Variations

The performance of current face recognition systems suffers heavily from the variations in lighting, especially under the outdoor environments [55]. The natural illumination in the real world (for example, the outdoor environment) is highly complex, consisting of reflected light from every direction as well as distributed and localized primary light sources. Recent public face recognition benchmarks demonstrated the seriously adverse effect of varying illumination in the performance of commercial face recognition systems. At the same time, the problem of coping with illumination variations is increasingly appreciated by the scientific community and significant progress has been achieved. Several techniques have been proposed in this area that may be roughly classified into two main categories.

The first category contains techniques seeking illumination insensitive representations of face images. Although it has been proven [56] that strictly speaking illumination invariants do not exist for Lambertian surfaces, several representations were seen to be relatively insensitive to illumination variability. For example, the direction of the image gradient was recently chosen to provide illumination insensitive representation and similarly the sum of gradient of ratios between probe and gallery images for an illumination insensitive similarity measure [57]. Savvides and Kumar [58] proposed the design of illumination insensitive correlation filters using a set of training images and varying illumination conditions. A similarity measure based on the difference between “prototype” images extracted from the corresponding probe and gallery images is proposed in [59] where the symmetry of the human face is also exploited. Recently, Shashua and Raviv [60] proposed the “Quotient image” as an illumination insensitive representation computed from the original image and a bootstrapping image set.

The second approach relies on the development of generative appearance models which is able to reconstruct novel gallery images resembling the illumination in the probe images. Exploiting the low dimensionality of the image space under varying illumination conditions and the Lambertian assumption [61], some of these techniques utilize a series of example images of the same person to reconstruct novel images [62]. Other approaches utilize a 3D range image and Albedo map of the person's face to render novel images under arbitrary illumination [63], while others are based on a combination of Lambertian reflectance [64] and harmonic images [65]. The shortcomings of this approach are: (a) the requirement in practice of large example sets to achieve good reconstructions, (b) the increase in the dimensionality of the classification problem by the introduction of the illumination variability, (c) the requirement of pixel wise alignment between probe and gallery images which necessitates pose compensation and (d) the reliance to simplified reflectance models (e.g. convex surfaces, shadow free, Lambertian reflectance).

2.9 Facial Tilt and Pose Variations

Drastic change of facial poses and tilt are big challenges for face recognition problems. Several techniques have been proposed to recognize faces under pose and facial tilt constraints. An active appearance model is proposed [66] using a generic face model iteratively fitted to the input image. The estimated control parameters (excluding those associated with pose) are subsequently used for face classification. The idea is extended by employing a high quality deformable model [67] generated from a large number of pixel aligned 3D scans and corresponding texture images. An analysis-by-synthesis technique is used to fit the model to the input image. Another approach is through the generation of novel views resembling arbitrary poses from a previously stored gallery image. Classification is subsequently based on the similarity between the probe image and the generated view. Beymer and Poggio [68] uses optical flow to estimate the deformation between the probe image and the corresponding frontal view gallery image. A face normalization procedure is described [69] based on the detection of facial features such as the eyes and mouth and adaptation of a generic 3D model. Various approaches to handling both illumination and pose variations are described by other researchers [70,71,72,73].

2.10 Image Resolution Variation

Image resolution is an important dimension to characterize face recognition performance. The change in image information that accompanies a change in image resolution mimics the information change caused by increasing viewing distances or common refractive errors in the optics of the eye. Understanding recognition under such conditions is of great ecological significance given their prevalence in the real world. What resolution face image is proper for face recognition? Simon Baker et al. [74] thought that the enhanced information of the high resolution over the low resolution face could be decided by the low resolution face and developed the corresponding face hallucination algorithm. Ce Liu et al. first constructed a recognized global model to hallucinate the individual global face, then built a local model to enhance the local face feature [75]. Face image resolution variation effects on recognition is still an open challenge to biometrics researchers.

CHAPTER 3

FACE IMAGES PRE-PROCESSING

3.1 Introduction

Face Recognition Technology (FRT) is a vital part of the broad area of pattern recognition. Face recognition is something that humans are exemplarily accurate at. Face recognition of faces with constraints like geometrical misalignment through facial tilt, noise, clutter and unnecessary information in image, and most importantly the varying image background require a basic set of tasks to be performed robustly. These all constraints are addressed in the preprocessing phase of recognition. Normalization includes the alignment and normalization of the face images and finally recognition is the representation and modeling of face images as identities. Few algorithms have been developed that automatically convert color images in to grayscale, carryout scale normalization, make the image background uniform and remove the facial tilt.

3.2 Color to Grayscale Conversion

It is possible to construct almost all visible colors by combining the three primary colors (i.e. red, green and blue) because the human eye has only three different color receptors, each of them sensible to one of the three colors. Different combinations in the stimulation of the receptors enable the human eye to distinguish approximately 350,000 colors. A Red, Green and Blue (RGB) color image is a multi-spectral image with one band for each color, thus producing a weighted combination of the three primary colors for each pixel.

A grayscale (or gray level) image is simply one in which the only colors are shades of gray. The reason for differentiating such images from any other sort of color image is that less information needs to be provided for each pixel. In fact a `gray' color is one in which the red, green and blue components all have equal intensity in RGB space, and so it is only necessary to specify a single intensity value for each pixel, as opposed to the three intensities needed to specify each pixel in a full color image.

Color images being in three planes of Hue, Saturation and Value are also computationally very extensive. To avoid handling of color images, they are converted to grayscale for faster processing. The most appropriate conversion method is by using expression 3.1 because it produces the result whose brightness is not equivalent to the brightness of original color image.

$$Y = 0.3R + 0.5G + 0.11B \tag{3.1}$$

$where, R = red, G = green, B = blue$

This function transforms a 24-bit, three-channel color image to an 8-bit, single-channel grayscale image. The explanation for these weights is due to the fact that for equal amounts of color the eye is most sensitive to green, then red, and then blue [76, 77]. This means that for equal amounts of green and blue light the green will nevertheless seem much brighter. Thus, the image obtained by the normal averaging of an image's three color components produces a grayscale brightness that is not perceptually equivalent to the brightness of the original color image.

3.3 Scale Normalization

Scale normalization has been handled through different methods which include stretching algorithm [78, 79] where locations of several feature points like eyes, nose, or mouth is used and unnecessary part of the image is discarded. In this research work, a new method to extricate the facial part from rest of image is used where face image is smoothen first by Gaussian convolution [80]. For smoothing process, the value of scaling parameter (sigma) which gives the width of filter is taken four. Increasing the width of Gaussian mask reduces the detector’s sensitivity to noise at the expense of losing some of finer details in the image. Experiments reflect that a tradeoff between filter sensitivity to noise and retention of desired finer details in the image is achieved by keeping the value of sigma four and in result an outer curvature of image is extracted.

$$F(i,j) = \frac{1}{2\pi\sigma^2} e^{-\frac{i^2+j^2}{2\sigma^2}} \tag{3.2}$$

Where *F(i,j)* is the image with dimension of *i, j* and σ is standard deviation of filter. To find the edge strength, a simple 2-D first derivative operator is applied on Gaussian smoothed image. Edges give rise to ridges in the gradient magnitude image. The algorithm then tracks along the top of these ridges and sets to zero all pixels that are not actually on the ridge top so as to give a thin line in the output- a process known as non-maximal suppression. The tracking process exhibits hysteresis controlled by two thresholds: TL=0.2 and TH=0.85. These hysteresis are used as a means of eliminating streaking. Streaking is the breaking up of an edge contour caused by the operator output fluctuating above and below the threshold. Here the aim is to extract the outer edge of the face image, so as to avoid undesirable and spurious edge fragments. Upper threshold is set quite high (i.e 0.85). At the same time, to eliminate the chances of noisy edges to breakup, lower threshold is kept too low (i.e 0.2). These values of thresholds help in extracting the outer curvature of the face as shown in Figure 3.1(c).

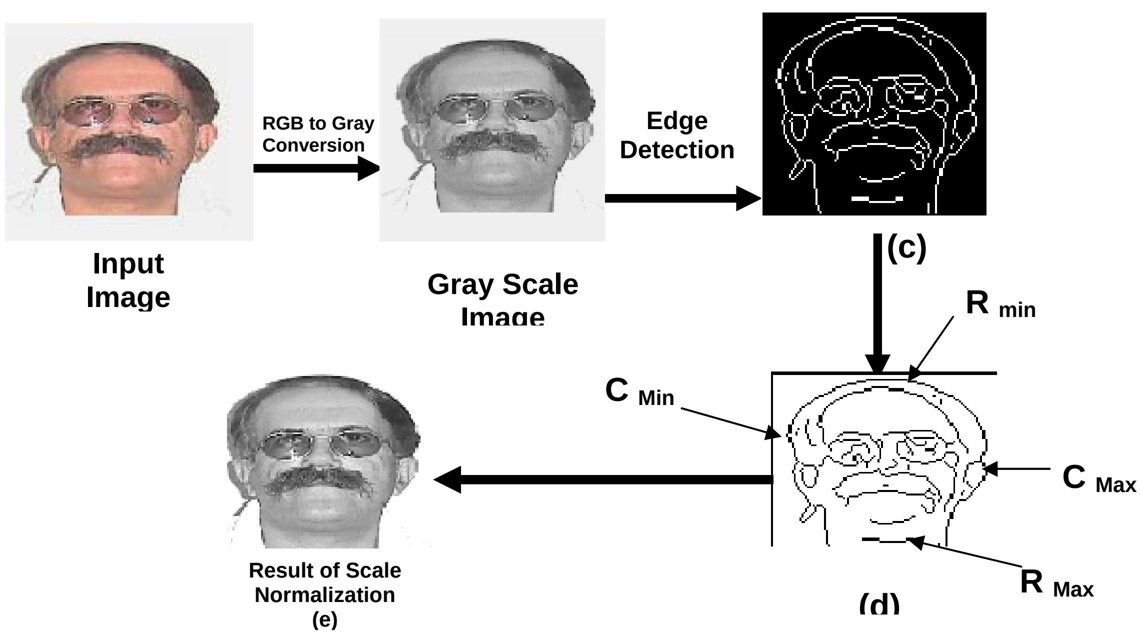


Figure 3.1. (a) Input Image (b) RGB to Grayscale Conversion (c) Controlled Edge detection (d) Classic Image Scan (e) Result of Scale Normalization

Binary image obtained in result of this edge detection is scanned from left to right, top to bottom in a classic pattern and four points shown in Figure 3.1(d) are worked out. The Image Scale Normalization (ISN) using the values of Equation 3.3 is carried out as shown in Figure 3.1 (e).

$$ISNF = ISNF_c\,, ISNF_R \tag{3.3}$$

$$where\ ISNF_c = C_{\max} - C_{\min}$$

$$ISNF_R = R_{\max} - R_{\min}$$

$ISNF$ = Image Scale Normalisation Factor

and $C_{\max}, C_{\min,}\ R_{\max}\ R_{\min}$ are maximum and minimum values of column and row respectively.

The normalized image is shown in Figure 3.1(e) which has been reconstructed by using calculated values of ISN.

3.4 Geometric Normalization

Geometric normalization is an important pre-processing step required in most of face recognition systems. An essential geometric normalization is removal of tilt in the facial images. As one of the salient features of the human face is eyes which can be considered salient and relatively stable feature of the face in comparison with other facial features. In addition, the size, location and the image-plane rotation of face in the image can be normalized by only the position of both eyes. Eye detection is divided into eye position detection and eye contour detection. However, most algorithms for eye contour detection require the detection of eye positions to initialize eye templates.

The first step is to calculate the gradient image of the face image [81]. Then the image is split into two halves. After this, a horizontal projection is applied to each of the two gradient images one by one. As we know that the eyes are located in the upper part of the face and the pixels near the eyes are more changeful in value comparing with the other parts of face. It is obvious that the peak of these horizontal projections in the upper parts of both images can give us the horizontal position of eyes. According to these horizontal positions and the total height of the face, we can easily line out the horizontal regions in both halves of the face in which the eyes are located. After the two rough regions of eyes are detected, template matching is used to locate the precise positions of iris centers in these regions. As the matching region reduces from the whole face to the two rough regions, the efficiency of algorithm is improved.

Suppose that we have a template T[x,y] according to the size of eyes of target image and we wish to detect its instances in target image I[x,y]. An obvious thing to do is to place the template at a location in an image and to detect its presence at that point by comparing intensity values in the template with the corresponding values in the image. Since it is rare that intensity values will match exactly, we require a measure of dissimilarity between the intensity values of the template and the corresponding values of the image.

In the case of template matching, this measure can be computed indirectly and computational cost can be reduced. We can simplify this measure by opening the square as given in Equation 3.4.

$$\sum_{|x,y|\in R} \left(I-T\right)^2 = \sum_{|x,y|\in R} I^2 + \sum_{|x,y|\in R} T^2 - 2\sum_{|x,y|\in R} TI \tag{3.4}$$

Now if we assume that I and T are fixed, then Σ *TI* gives a measure of mismatch. A reasonable strategy for obtaining all locations and instances of the template is to shift the template and use the match measure at every point in the image. Thus, for an m×n template, we compute best match *E* [*x*,*y*] by using Equation 3.5.

$$E\left[x,y\right] = \sum_{s=1}^{m}\sum_{d=1}^{n} T\left[s,d\right] I\left[x+s,y+d\right] \tag{3.5}$$

where s and d are the displacements with respect to the template in the image. This operation is called the cross-correlation between T and I. The aim is to find the locations that are local maxima or are above a certain threshold value. However, a minor problem in the above computation was introduced when we assumed that T and I are constant.

When applying this computation to images, the template T is constant, but the value of I will be varying. The value of E will then depend on I and hence will not give a correct indication of the match at different locations. Normalized cross-correlation (C_{TI}) can solve this problem. The match measure E then can be computed using Equation 3.6.

$$C_{TI}\left[x,y\right]=\sum_{s=1}^{m}\sum_{d=1}^{n}T\left[s,d\right]I\left[x+s,y+d\right]$$
$$E\left[x,y\right]=\frac{C_{TI}\left[x,y\right]}{\left\{\sum_{s=1}^{m}\sum_{d=1}^{n}I^{2}\left[x+s,y+d\right]\right\}^{\frac{1}{2}}} \tag{3.6}$$

The local maxima in the above computation gives the position of the iris centers. Let us say that they are the two points (xl,yl) and (xr,yr) for the right and the left eye respectively. These are then used to compute the tilt (slope m and angle θ) in the image using Equation 3.7 and 3.8.

$$m = (yr - yl) / (xr - xl) \tag{3.7}$$

$$\theta=\arctan(m) \tag{3.8}$$

Finally, the tilt is removed using the reverse rotation (i.e., rotating by –θ) as shown in Figure 3.2.

Figure. 3.2. Image with and without Tilt

3.5 Varying Image Background Issues in Face Recognition

Varying background of images contribute to failure rate of pattern recognition techniques. Since a face is elliptical in shape, no matter how tight a bounded region is determined, with a rectangular box, there are bound to be sections of the background included inside the face image. This issue can be settled in two ways either to remove the background or to make it uniform throughout the face image. These methods have been applied in this work.

3.5.1 Uniform Image Background

Mathematical morphology (MM) emerges as a general theory that has provided unified approach to deal with problems in image segmentation [82]. MM also provides operations which simplify images preserving their intrinsic shape characteristics [83]. To minimize image background effects on face recognition, grayscale morphological operations with a disk type structuring element are used. Gray level opening operation is applied on face images which is nothing but erosion followed by dilation where erosion tends to shrink the white regions of an image and dilation tends to grow the white regions of an image.

Erosion is an operation where the structuring element values are subtracted from the image during the scanning process and a minimum of each set of values is returned. This may be implemented as a correlation-like operation where the extremum operator Min and subtraction replace multiplication and addition respectively. For a (2M+1) × (2M+1) structuring element B and an N×N image array A, the erosion of A by B may be defined as:

$$C_e(n_1,n_2) = \min_{i,j}\left\{A(n_1+i,n_2+j) - B(i,j)\right\}$$
$$-M \le i,j \le M,\ \text{with } 0 \le n_1,n_2 \le N-1 \tag{3.9}$$

Grayscale dilation is a convolution-like operation where a typically structuring element is scanned over an image and at each position, the maximum of point-by-point sums of the image and the structuring element is computed. The dilation of A by B may be defined as:

$$C_d\left(n_1,n_2\right)=\max_{i,j}\left\{A\left(n_1-i,n_2-j\right)+B\left(i,j\right)\right\}$$
$$-M\le i,j\le M,\ \text{with } 0\le n_1,n_2\le N-1 \tag{3.10}$$

The disk type structuring element is used to preserve foreground region that have a shape similar to it or that can completely contain the structuring element. Later on, the background image obtained through this process is subtracted from the original image to minimize the varying background image contribution for recognition.

3.5.2 Image background Removal

To minimize its influence on failure rate of pattern recognition, it is made uniform through image segmentation. Eight bit gray scale images are converted into three bit to reduce the gray scale variation within background regions and later on Median filter of size n=5 is applied on the image. Median filter forces points with distinct gray levels to be more like their neighbours. Isolated clusters of pixels that are light or dark with respect to their neighbours and whose area is less than $n^2/2$ are also forced to median intensity. Afterwards image background with low range gray scale values is addressed on region based approach and changed into single value in corresponding original image to make it uniform through out the dataset. Segmented and final images are shown in Figure 3.3.

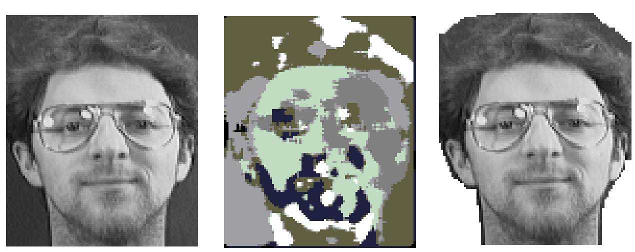


Figure.3.3. (From left to right) Original, segmented and final Image

3.6 Illumination Variation Control

Illumination is another recognition constraint which drastically affects the recognition results. Non-uniform illumination condition during acquisition process play a significant role in recognition methodology even if image background is removed as illumination effects change the facial representation. Illumination variation can be divided into following four categories as shown in Figure 3.4.

- Normal illumination condition
- Uniform illumination with linear scale of lighting intensity
- Non-uniform illumination (light source may be from different directions)
- Uniform illumination with nonlinear scale of lighting

Figure 3.4. Examples of Illumination Variations

The second case can be adjusted by normalizing the whole image intensity. If $f(n, m)$ represents the face image and the normalized face image is then represented by $I(n, m)$ which is defined as:

$$I\left(n,m\right)=\frac{f\left(n,m\right)}{\|f\|} \tag{3.11}$$

Where

$$\|f\|=\sqrt{\sum_{n=0}^{N}\sum_{m=0}^{M}f\left(n,m\right)^{2}} \tag{3.12}$$

In the fourth case, if the probe image is in uniform illumination with non-linear scale of lighting intensity, the gray-level intensity of the image is adjusted using gray-level histogram equalization. The most difficult situation is the third case in which the light source may come from different directions. Basically, it is hard to find a generic model to solve this case. The effects can be minimized through background removal.

3.7 Databases used to Evaluate the Recognition Models

Most of the face recognition techniques, due to the statistical nature of the problem, are dataset dependent. A technique might perform better under a given set of conditions and may perform poorly for another set of conditions. The performance of face recognition algorithms usually degrades as more subjects are added to the database, due to the increasing probability of the presence of subjects with similar attributes. Following different datasets are used for experiments:

3.7.1 ORL Database

The ORL database contains photographs of faces taken between April 1992 and April 1994 at Olivetti Research Laboratory (ORL) in Cambridge, UK [84]. It consists of 400 images of 40 individuals, 10 images were taken for each individual and few constraints on facial expression and pose were imposed. Furthermore, some of captured images were subjected to illumination variations. This database contains both male and female images and it is expected that this is a more difficult database to work with. There are variations in facial expressions (open / closed eyes, smiling / non-smiling) and facial details (glasses / no glasses). There is some variation in the scale up to about 10%. An example of variation in the images is shown in Figure 3.5.

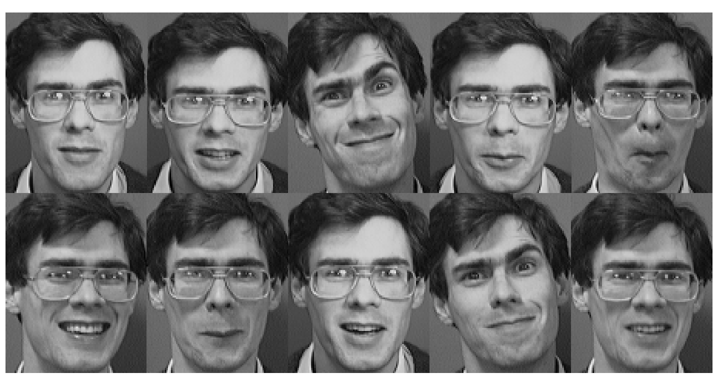


Figure.3.5. Examples of ORL Dataset Images

3.7.2 Yale Database

The YALE database contains 165 gray scale images in GIF format of 15 individuals [85]. There are 11 images per person, one per different facial expression or configurations: center-light, with or without glasses, sad, happy, sleepy, surprise and wink. Some examples are shown in Figure 3.6.

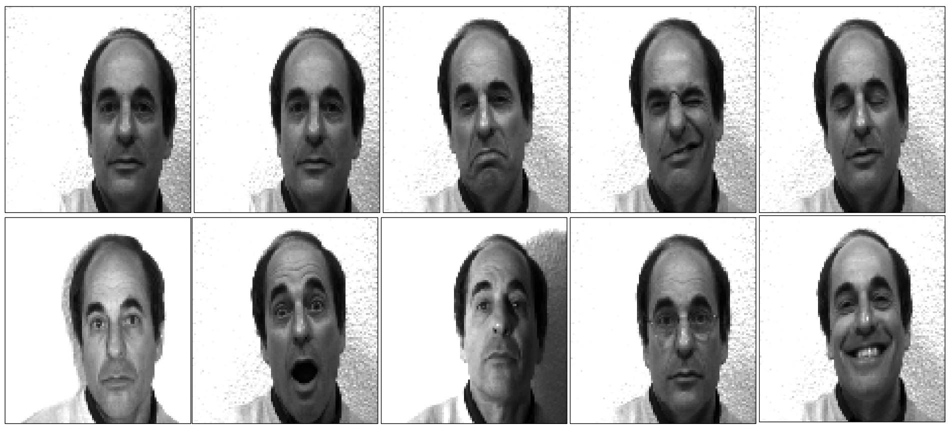


Figure.3.6. Examples of Yale Dataset Images

3.7.3 CEME_ NUST Color Database

College of Electrical & Mechanical Engineering _ National University of Sciences and Technology (CEME_NUST) color dataset consists of 15 sets of color images [86] of different individuals with 10 varying poses, sizes, illuminations. Different facial expressions and occlusions were taken at image processing lab of College of E&ME, National University of Sciences and Technology Rawalpindi, Pakistan. This dataset is part of this research work and particularly obtained for testing of the developed face recognition methods on color images. Few examples of this data set are shown in Figure 3.7.

Figure.3.7. Examples of CEME_NUST Color Dataset Images

3.7.4 FERET Database

The FERET program set out to establish a large database of facial images that was gathered independently from the algorithm developers. Dr. Harry Wechsler at George Mason University was selected to direct the collection of this database. The database collection was a collaborative effort between Dr. Wechsler and Dr. Phillips. The images were collected in a semi-controlled environment. To maintain a degree of consistency throughout the database, the same physical setup was used in each photography session. Because the equipment had to be reassembled for each session, there was minor variation in images collected on different dates. The FERET database was collected in 15 sessions between August 1993 and July 1996. The database contains 1564 sets of images for a total of 14,126 images that includes 1199 individuals and 365 duplicate sets of images [87]. A duplicate set is a second set of images of a person already in the database and was usually taken on a different day. For some individuals, over two years had elapsed between their first and last sittings, with some subjects being photographed multiple times. This time lapse was important because it enabled researchers to study, for the first time, changes in a subject's appearance that occur over a year. All images are of 256 x 384 pixels size. We have taken 10 images of 100 persons with a total of 1000 images for our experiments. Figure 3.8 shows few examples.

Figure.3.8. Examples of FERET Dataset Images

3.7.5 CMU AMP Face Expression Database

This dataset consists of 13 subjects with 75 images each. All images were collected under same lighting conditions and only facial expressions for each image were allowed to vary [*88*]. This dataset provides adequate number of images with varying facial expressions for evaluation of effects of expression changes on recognition rate. Few examples of varying expression are shown in Figure 3.9.

Figure 3.9. Examples of CMU AMP Face Expression Database

CHAPTER 4

FACE IMAGE DIMENSION REDUCTION THROUGH IMAGE RESOLUTION VARIATION

4.1 Introduction

Digital images are made up of pixels. Pixels are the small sections of color and/or tone that together form a digital image. Pixels form an image-like pieces of a mosaic. A digital image is a grid of pixels and if the pixels are viewed together in proper registration, the image is formed. When there are enough pixels and they are small enough so as not to be individually discernible, the digital image can achieve photo quality. A digital image as it is created by a digital camera or scanner and as it resides on digital storage medium like a hard drive or a flash card is simply an informational record of a grid of pixels - a map of specific tones at specific locations (bitmap). It is common to refer to computer records as files, and thus digital images in this state, and generally, are commonly referred to as "digital image files," "image files," or even just "files." The image has no physical size, but if stored in certain file formats the image may have a specific size designated as part of the informational record, this can easily be set or later adjusted to any size. When an image is displayed or output it takes on physical form as an image. Whether being displayed on a monitor or printed on paper, the physical image has spatial dimensions - width and height. The same digital image can assume many different physical sizes.

Resolution refers to the number of pixels in an image. Resolution is sometimes identified by the width and height of the image as well as the total number of pixels in the image. Image size, resolution and compression are interrelated terminologies. Sometimes an image cannot be reproduced at the size needed. It lacks the needed resolution or has too much for practical use. In these cases, the computer can be called upon to modify the resolution of an image, reducing its resolution or increasing it. Different techniques to vary the image resolution are employed which have effects not only on quality of image but also on image recognition. This is demonstrated by evaluating effects of varying image resolution using different techniques of face recognition.

4.2 Types of Image Resolution

Digital image resolution designations fall into two basic categories: Pixel Count Resolution and Spatial Resolution.

4.2.1 Pixel Count Resolution

Pixel count resolution is simply the amount of pixels a digital image contains, or made up of. It is expressed in either mega pixels or pixel dimensions. The term mega pixels simply mean millions of pixels. A six mega pixel image is an image made up of 6 million pixels, or some rough approximation thereof. Pixel dimensions represent another more descriptive method of designating pixel count resolution. By stating the pixel dimensions of a digital image, the pixel count is inferred. In addition, the aspect ratio of the image is revealed as are the exact dimensions. Pixel count resolution is a fixed property of an image. Unless the image is resampled or cropped, the image remains with the same number of pixels.

4.2.2 Spatial Resolution

It relates to the number of pixels in a spatial measurement of a physical image - pixels per inch. Spatial resolution does not apply to an image file (except as a temporary/variable specification thereof), only to a physical image. An image literally can not have a spatial resolution if it doesn't take a physical form - it can't have any given number of pixels per inch if it doesn't have physical dimensions. The spatial resolution of an image is commonly referred to in terms of dots per inch (dpi). What is being specified is pixels per inch, however "dots" per inch has gained a foothold in common terminology. Spatial Resolution is a variable property of an image file - it only becomes a fixed property of an image once it is output in permanent form (i.e. printed). As this resolution is conditional upon output, this resolution is commonly called output resolution or print resolution.

4.3 Image Resolution and Face Recognition

In case of face recognition, behavior of facial images changes with variations in image resolution. Resolution variation is directly connected with the information possesses by face images. This information is of two types (i.e. discriminative information and structure information). Discriminative information represents the information distinguishing from other face images whereas structure information represents the common information of all the face images under the same resolution, which can be roughly estimated from low resolution face image. Similarly difference of the discriminate information between higher resolution face image and the lower face image is called enhanced discriminative information and difference between structure information is called enhanced structure information. When observing an object, people will usually see it more clearly when moving nearer and see it most clearly when distance is smaller than one fixed value L_o. In case of face image, we recognize it when distance is another fixed value L_1 which is usually larger than L_o. This conjecture can be supported through Shanon sampling theory which describes that face image with one certain resolution can retain all the original image information [89] (i.e., the corresponding reconstruction image will not lose any information). Similarly, the central challenge in face recognition lies in understanding the role of different facial features play in varying image resolution. These features can be the internal (eyes, nose and mouth) and external (hair and jaw-line) features. Most important is the relative contributions of internal and external features change as a function of image resolution. One very interesting conclusion on issue is that reports from several researchers studying face recognition by neonates [90] suggest that infants initially depend more on external features than on internal ones for discriminating between individuals, external features are more useful in face recognition when image resolution varies through decimation. Visual system uses a highly non-linear cue-fusion strategy in combining internal and external features along the dimension of image resolution and that the configural cues that relate the two feature sets play an important role in judgments of facial identity.

4.4 Image Resolution Variation Through Interpolation

Image resolution variation is always required in image processing operations as sometimes image under consideration lacks the needed resolution or has too much for practical use like recognition purposes. In these cases, various techniques are employed to alter the image resolution. Interpolation and image decimation are two distinct methods to achieve the image resolution variation.

Interpolation is estimation of an image pixel value at a location in between image pixels. The process is called into force when there is not enough resolution in an image to accomplish the reproduction needed. Interpolation can be a daunting task for a computer, in which it carries out the analysis and averaging the new pixel values and colors that are inserted in an image. The process of interpolation requires the computer to measure adjacent pixels and "guess" a correct value for new pixels inserted in the data stream. In the process, it can make errors, create unwanted noise, and introduce abnormal pixels into areas of the image causing "artifacts." The process is good enough, though, to produce an acceptable image in most cases. The important thing to remember is that:

$$\begin{bmatrix} \alpha > 2 \text{..........Resultant image quality is unacceptable} \\ \alpha = 2 \text{..........Resultant image quality is acceptable} \end{bmatrix} \tag{4.1}$$

where α is the interpolating factor that assume positive discrete values only α =1 implies no interpolation α =2 reflects that size of interpolated image is double the size of original image.

Any interpolation is potentially harmful to an image and some interpolation may ruin it. If an image is interpolated to a small degree (up to a doubling of the image's original size), this is usually harmless. Beyond two times the original size, you will begin to see serious effects in most images. The types of interpolation are:

- **Nearest-Neighbor Interpolation.** Output pixel values are assigned the value of the pixel that the point falls within. No other pixels are considered.

- **Bilinear Interpolation.** Output pixel values are calculated from a weighted average of pixels in the nearest 2-by-2 neighborhood.
- **Bicubic Interpolation** Output pixel values are calculated from a weighted average of pixels in the nearest 4-by-4 neighborhood.

4.5 Image Gaussian Pyramid through Decimation

In most pattern recognition systems, image resolution reduction can be achieved through image decimation. Decimation algorithm is a novel form to obtain desired resolution where the sub-sampling under the best circumstances is harmless to image as it scans through lines of pixels, averaging together pair of pixels or group of pixels according to the value of Decimation factor (G). The resulting image is a reduced size mirror of the original image faithful in tonality to the original image but smaller in size. G can be calculated as:

$$G = P / Q \tag{4.2}$$

where

P = Order of original image matrix

Q = Order of desired decimated image matrix

G = Arbitrary down scale decimation factor

The sliding mask shown in Figure 4.1 is applied on face images for the decimation. The simultaneous process of low pass filtering and decimation is performed where a low pass averaging filter is applied on face images and every next pixel according to order of 2D filter is decimated. The mathematical model of decimation process is as under where O(i,j) is original image and M,N is order of resulting image:

$$I(M,N)=\frac{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left[\sum_{m=i\times\mu_w}^{i\times\mu_w+(\mu_w-1)}\sum_{n=j\times\mu_h}^{j\times\mu_h+(\mu_h-1)}O_{i\times j}\right]}{G\times G} \tag{4.3}$$

$$\mu_w=\frac{W}{D_w} \quad \& \quad =\mu_h=\frac{H}{D_h}$$

Where

W & H = Width and Height of the original image

D_w & D_h = Width and Height of desired output image

μ_w = Ratio between actual width of image and desired width

μ_h = Ratio between actual height of image and desired height

$$\frac{1}{9}\begin{array}{|c|c|c|}\hline \mathbf{1} & \mathbf{1} & \mathbf{1}\\ \hline \mathbf{1} & \mathbf{1} & \mathbf{1}\\ \hline \mathbf{1} & \mathbf{1} & \mathbf{1}\\ \hline\end{array}$$

Figure 4.1. Decimation mask of 3 x 3

The sliding mask scans through the whole image and averages together the pixels under it. In scanning process, factor G determines the jump of the mask. The order of the mask determines the G and accordingly the resolution is varied. By changing the value of G, a Gaussian Pyramid of varying image resolution as shown in Figure 4.2 is obtained. This decimation algorithm is applied on all the images of database by varying the value of G and recognition results are obtained against these images with varying resolution.

4.6 Effects of Varying Resolution on Template Matching Face Recognition Model

Nearest-Neighbor template matching is the most native method of face recognition [91,92], where individuals are represented by a set of complete images of their face. An input face is compared with each image in the database. The image for which the template matching gives the highest response is considered to be the match. Proposed template matching model of face recognition has been modified where images were centered first and then through decimation process image resolution is varied. Proposed model of face recognition is shown in Figure 4.3.

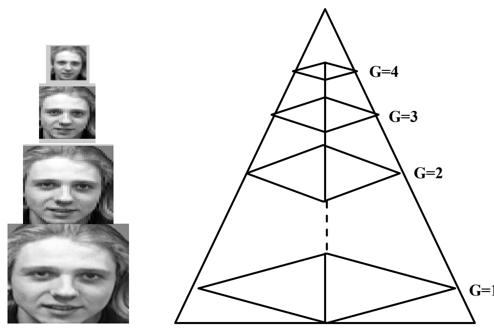


Figure 4.2. Gaussian pyramid

4.7 Resolution and Dimension Reduction through Decimation Algorithm

Decimation algorithm is applied on each face image of database. By varying the value of G, the resolution of each image is changed and a Gaussian pyramid of each image is obtained. In learning process of proposed face recognition model, five images of each class at one particular resolution are taken and their mean $\overline{M}$ is calculated by:

$$\overline{M}_{mxn} = \frac{\sum_{i=1}^{T} I_i}{T} \tag{4.4}$$

Where:

I_i is the ith image with specific Gaussian base value used for training of the system

T varies from one to total number of images used for training

Images are centered by subtracting mean from each image. It produces a dataset whose mean is zero.

$$\underset{m\times n}{C(T)} = \left[\underset{m\times n}{I(T)} - \overline{M}_{m\times n} \right] \tag{4.5}$$

The subspace of centered images is stored in the system as the reference against each individual image of training set.

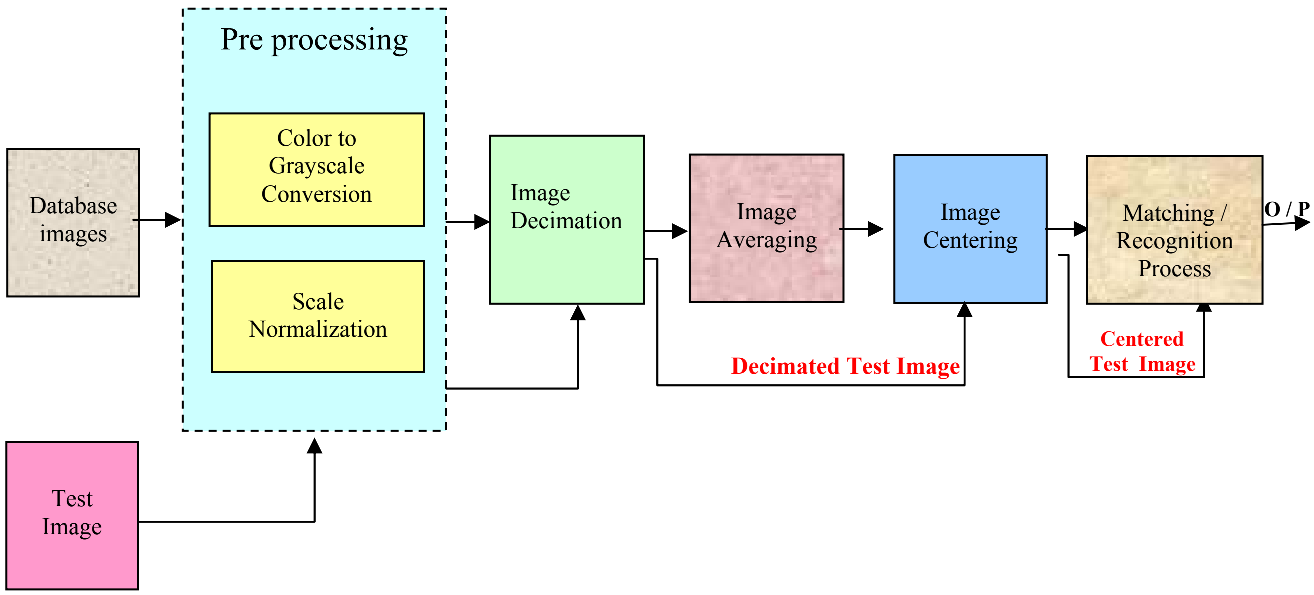


Figure 4.3. Model of the System

In testing phase dimension of preprocessed test image is reduced as done in training of model and then it is centered by subtracting mean from it. For image classification, Euclidean distance is used as matching criterion. This process is repeated for next level of resolution of same database until Face Recognition Rate (FRR) on all the images of Gaussian pyramid is obtained.

4.8 Dissimilarity Space and Matching

In training of the model, five images of each subject are used and decimated image matrix of each image used in training process is obtained.

$$I^i = \left[I_1^i \, I_2^i \cdots\cdots \, I_N^i \right]^T \tag{4.6}$$

In matching process, a dissimilarity space *D* (I_i, *P*) of test image with training images is obtained by using simple Euclidean distance. This dissimilarity space matrix:

$$E = D\left(I_i, P\right) \tag{4.7}$$

is converted to a vector and the vector with min difference is taken as matched image:-

$$R = arg\ min\left[E\right] \tag{4.8}$$

where I is training image and $i = 1, 2, \ldots\ldots\ldots, total\ subjects$ used in training of model, *R* is the recognized image and E is the difference matrix of test image with all training images. Here the $arg\ min$ which is accumulated minimum Euclidean distance provides the best match.

4.9 Implementation of Face Recognition Model

This model of face recognition has been implemented by developing following modules and data structures.

4.9.1 Scale Normalization Module

In this module, training images of database are loaded in an array. Dimension of the images is checked to distinguish between color and grayscale images. If image dimension is three (i.e. red, green and blue), images are converted to grayscale. Gaussian filter is applied on grayscale images to smooth the images. In next step, the edge detection through 2D image derivate is obtained. Upper and lower threshold values (i.e 0.85 and 0.2) are used to obtain the thin outer curvature of face image. Resultant image is scanned in a loop from left to right and from top to bottom to find out expansion of binary image in up, down, left and right directions. Four outer most values are calculated in result of this scanning process. Finally the original input image is rescaled according to calculated values. The flowchart of scale normalization module is shown in Figure 4.4.

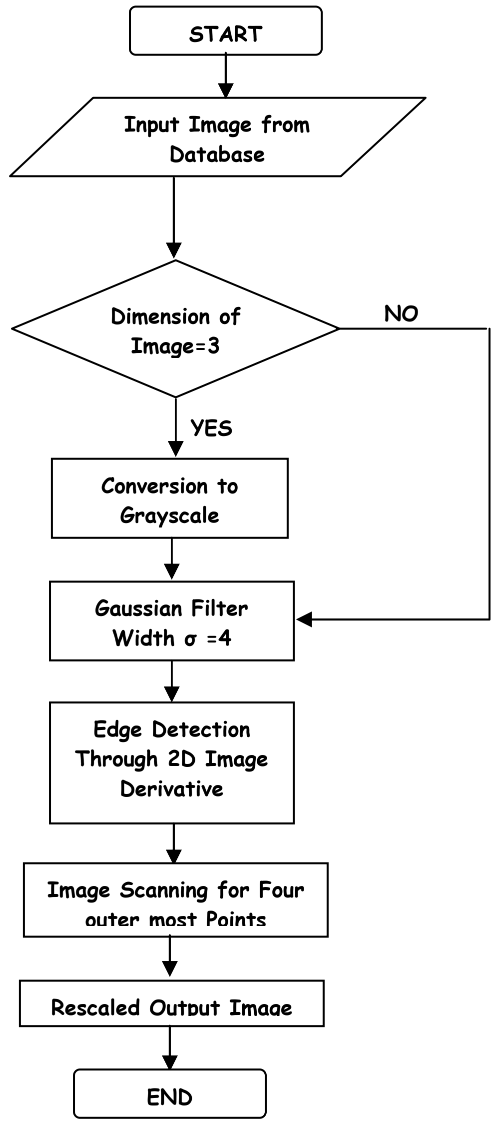


Figure.4.4 Flow chart of Scale Normalization Module

4.9.2 Geometric Normalization Module

Geometric deformation (mainly facial tilt) is calculated and removed in this module. Scale normalized grayscale image is loaded as input to this module. Following algorithm has been developed and implemented to remove the facial tilt.

Begin

1. Repeat till end of training images

2. Carryout eye localization

- Load scale normalized grayscale image
- Calculate gradient of image
- Find out horizontal projection
- Calculate horizontal line of eyes
- Find vertical projection
- Calculate two peaks
- Determine size of eyes

3. Carry out eye template selection

- Load eye templates of different sizes
- Make selection of template according to face image eye size

4. Obtain iris localization

- Carry out scanning of template image over face image
- Obtain cross correlation
- Locate iris of both eyes

5. Carry out tilt compensation

- Compute slope and angle between two eyes
- Apply reverse rotation

End

To remove the facial tilt, first rough region of both eyes is located through feature based method. Precise iris location of both eyes is determined by using template matching in areas of both eyes. Slope and angle between two eyes give amount of facial tilt present in the image which is finally compensated by applying reverse rotation.

4.9.3 Background Module

After scale and geometric normalization, background module has been employed to make the image background uniform by using grayscale morphological operations, where each pixel in output image is based on a comparison of the corresponding pixel in the input image with its neighbours. A disk type kernel (which conserves the foreground area that has the shape like it) is used to erode and dilate the image. It is termed as structuring element. First, image is eroded by scanning the structuring element over the image. The values of structuring element are subtracted from the image values and $\arg Min$ is used to return the final value. In next step, eroded image is applied as input to the dilation process which is just a convolution like operation. Structuring element is scanned over the image and pixel by pixel sum with the structuring element is computed and $\arg Max$ is used to return the value. This dilated image is subtracted from input image which gives uniform background as compared to original image.

4.9.4 Resolution Module

This module is devised to vary the image resolution according to the decimation down scale factor. Preprocessed images used for training of the system are provided to convolution sub module which scans a low pass averaging filter (shown in Figure 4.1) over the entire image. Convolution process is carried out according to Equation 4.9 which returns the averaged value of all pixels under the filter window at that particular instance.

$$I(m,n) = O(i,j) * h(x,y) \tag{4.9}$$

$$I(m,n) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} O(i,j)\, h(x-i, y-j) \tag{4.10}$$

Convolved image is applied to image decimator which throws away the every next pixel according to decimation down scale factor. Resultant image is a condensed size image with reduced resolution. By varying the decimation rate, images of varying resolution are obtained. An image with original resolution and one with reduced resolution through image decimation is shown in Figure 4.5.

Figure. 4.5. Original and Decimated Image (left to right)

4.9.4 Training Module

Training part of the proposed face recognition model is implemented in this module where five preprocessed and decimated images of each class are loaded as input. First mean of these images is calculated and every image is centered by subtracting the mean from it. Decimated images provide dimension reduction and at the same time reduce the variation within class images. These images with reduced dimension are converted into a column vector. A subspace of dimension of L x T is attained whose each column represents an image. Where L is total number of pixels of decimated image and T is total images used for training of the model.

4.9.5 Recognition Module

Recognition module is last module of the proposed face recognition model where a test image is presented as input. Test image is first preprocessed and afterwards it is decimated with same value of down scale factor as used in training module. Then image is converted into a column vector with a dimension of L x 1. In matching process, the Euclidean distance of this vector with each column of subspace gathered during training of model is acquired and finally a difference matrix of dimension L x T is obtained.

Euclidean distance of each image is accumulated, a matrix of L x T is converted into 1 x T dimension and *arg Min* is used to find out the best match.

4.10 Experiments and Results

Four different databases (ORL, YALE, FERET and CEME_NUST color database) as discussed in chapter 3 have been used to obtain the Face Recognition Rate (FRR) against varying image resolution. Figure 4.6 illustrates the results of ORL and YALE databases. Images in both the databases were acquired at resolution of 112 x 92 and tested on various resolution levels of 112 x 92, 56 x 46, 28 x 23, 18 x 15 and 14 x 11. Images of CEME_NUST color database and FERET have been acquired at resolution of 112x100 and 256x384 respectively. CEME_NUST color database has been tested on varying image resolutions (112x100, 56x50, 28x25, 18x16 and 14x12) and FERET images have been evaluated at different resolution levels (256x386, 128x193, 64x96, 42x64 and 32x48). Plot lines of Figure 4.6 and Figure 4.7 show that as the image resolution of face images is reduced FRR is improved and reaches at optimum value at a specific image resolution. Above or below this specific resolution FRR is influenced negatively due to inclusion of redundant data or lack of required facial information for recognition in image used for recognition.

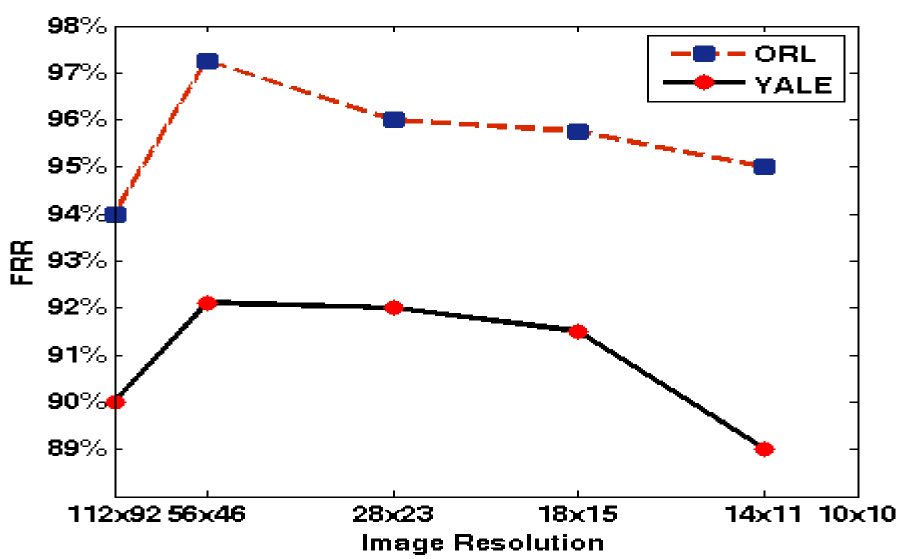


Figure.4.6 FRR of ORL and Yale databases with varying image resolution

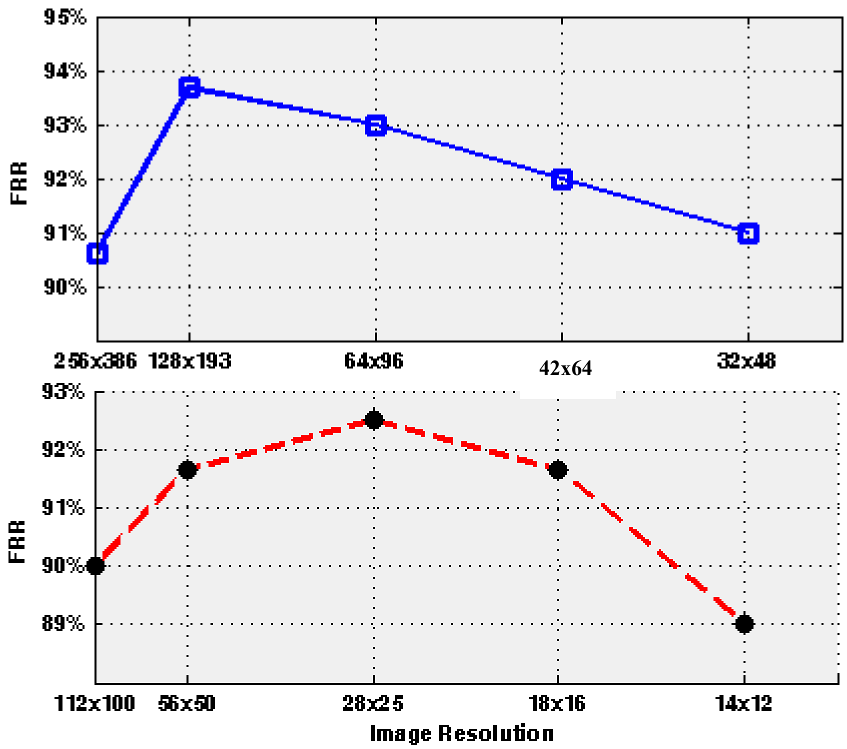


Figure.4.7 FRR of FERET database (upper) and CEME_NUST Color database (lower) with varying image resolution

The image resolution against optimum FRR varies from database to database as ORL database images provides FRR of 97.2% at resolution of 56 x 46, Yale database images provide FRR of 92% at resolution of 56x46, CEME_NUST color database provides FRR of 92.5% at resolution of 28 x 25 and FERET database showed optimum performance of 93.7% at resolution of 128x193.

Reduced image resolution not only provides improved FRR but also addresses the curse of dimensionality with improved speed. Resolution variation through image decimation presents a considerable image dimension reduction which ultimately reduces the database training time and image recognition time. Table 4.1 reflects the database dimensions against which best FRR is achieved and correspondingly the reduction in training and recognition time are encircled. As the image resolution is reduced accordingly image dimensions are reduced which improves the computation speed of the recognition model. Best FRR is obtained in ORL and YALE database at dimension reduction of 56x46 from 112x92, CEME_NUST Color database provides optimum FRR at image dimension of

28x25 which is four times image resolution reduction whereas FERET database offered best performance at 128x193 which is twice image resolution reduction.

Table.4.1 Training and Recognition time against Varying Image Resolution

Image Resolution	**Training Time (Seconds)**		**Recognition Time per Image (Seconds)**		**ORL Database**	**Yale Database**
	ORL Database	**Yale Database**	**ORL Database**	**Yale Database**	**FRR (%)**	**FRR (%)**
112x92	28.78	9.25	0.217	0.1048	94	90.0
56x46	**11.47**	**3.03**	**0.111**	**0.053**	**97.25**	**92.12**
28x23	5.337	1.79	0.072	0.033	96.0	92.00
18x15	3.585	1.4	0.064	0.029	95.75	91.50
14x11	3.175	1.32	0.060	0.026	95	89.00

4.11 Discussion of Results

Template matching is a very native technique of face recognition which has been implemented to evaluate the changes in recognition rate caused by image resolution variation. As reflected from Figures 4.6 and 4.7, for each database recognition rate changes with change in image resolution up to a certain level. This conjecture supports the phenomena that in each dataset up to certain image resolution, the image information is not significantly lost. Moreover, in decimation process by applying the averaging filter high frequency components present in the image are reduced while retaining the low frequency components intact which may reconstruct the image without any significant loss to the image quality. This reduction of undesired high frequency components for recognition not only minimize the redundant clutter of facial information but also improve the computation time of model and provides better matching as it is analytically proved in Tables 4.2 & 4.3. In Table 4.2 first test image is S12-6 of CEME_NUST database which means sixth image of S12 class. Once its resolution is set to 112x100 or 56x50, its best matching is with second image of S11 class with Euclidean distance of 3.1284 x 10^6, hence giving wrong recognition results. Once its resolution is further reduced to 28x25 Table 4.2 reflects that its best matching is with S12-1 having Euclidean distance of 6.364x 10^5 which is first image of its own class, so resulting in correct recognition. Similarly S6-9 and S8-7 images of same database have been also used as test images and results are illustrated in Table 4.2.

Table.4.2. Results of CEME_NUST database test images at different values of "G"

Test Image	G	Best Euclidean Distance within class	Best Match Image within class	Best Euclidean Distance out of class	Best Match Image out of class
S12-6	2	3.174 x 10^6	S12-1	3.1284 x 10^6	S11-2
S12-6	4	6.364x 10^5	S12-1	Correct Match with own Class	-
S6-9	4	1.696x 10^6	S6-3	Correct Match with own Class	-
S6-9	8	3.762 x 10^5	S6-3	3.362x 10^5	S5-4
S8-7	4	1.876 x 10^6	S8-1	Correct Match with own Class	-
S8-7	8	4.117x 10^5	S8-1	3.4766 x 10^5	S9-3

The conjecture, "face recognition behavior of face image changes with change in image resolution" has also been analyzed by using images of ORL database at different resolutions. Table 4.3 and Figure 4.8 reflect that when S1-10 test image with resolution of 112x92 is presented to model it provides minimum matching difference with image of wrong class i,e. S17-2. Once same image is offered at reduced resolution of 56x46, it provide correct recognition with minimum matching difference with S1-2 of its own class. If resolution is further reduced Figure 4.8 reflects that again false recognition with S17-5 is obtained.

These results conclude that at certain resolution facial feature of face images become so promising to the facial feature of training images that it gives better recognition results as compared to results obtained at normal resolution. As each database images are acquired at different image resolution so every database provides improved FRR at different resolution. Table 4.4 reveals the best image resolution of each database against optimum FRR.

Table.4.3. Test image results of ORL database with different values of "G"

Test Image	**G**	**Best Euclidean Distance within class**	**Best Match Image within class**	**Best Euclidean Distance out of class**	**Best Match Image out of class**
S1-10	2	1.215×10^{6}	S1-5	Correct Match with own Class	-
S1-10	4	2.971×10^{6}	S1-4	2.954 x 106	S17-5
S3-8	2	8.154×10^{6}	S3-1	Correct Match with own Class	-
S3-8	6	7.453×10^{5}	S3-1	7.302 x 105	S38-4
S27-8	2	8.343×10^{6}	S27-5	Correct Match with own Class	-
S27-8	8	4.321×10^{5}	S27-5	4.030 x 105	S4-1

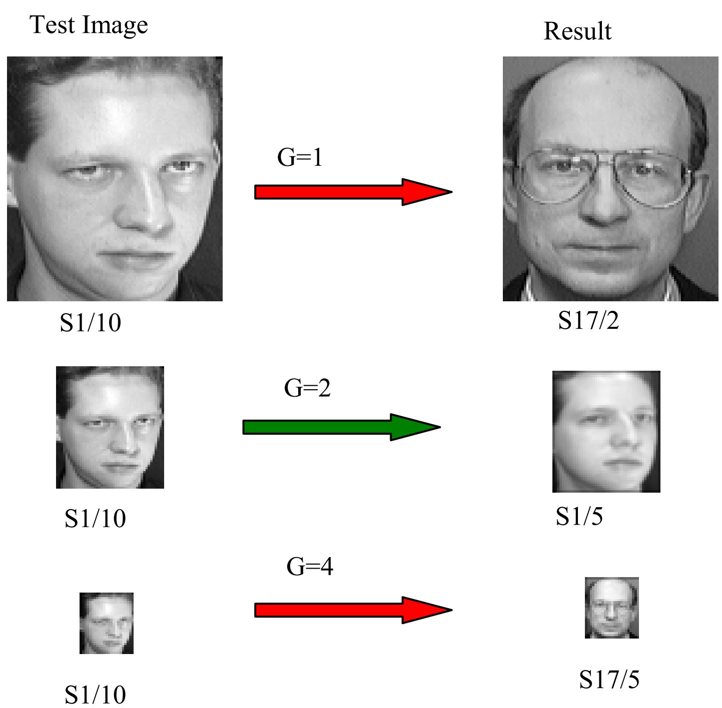


Figure 4.8. Effects of Changing Resolution on Recognition

Disparity of image resolution against best FRR among these databases is due to two factors:- (1) different database images have different original resolution as ORL and Yale database images have resolution of 112 x 92, where as EME color database images have resolution of resolution of 112 x 100 and FERET database images are obtained at resolution of 128 x 192 and (2) strength of high frequency components present in images differ from database to database. These two factors contribute towards best FFR at different resolutions in different databases.

Table.4.4Image Resolution against optimum FRR

Name of Database	**Image Resolution against Best FRR**	**FRR (%)**
ORL	56x46	97.2
YALE	56x46	92
FERET	128x193	93.7
CEME_NUST	28x25	92.5

These results also proved that change in image resolution by a certain factor in each dataset removes the unnecessary clutter of information (i.e high frequency components) from image which is not required for the face recognition and at the same time achieve considerable dimension reduction.

4.12 Principal Component Analysis

Principal Components Analysis (PCA) is a way of identifying patterns in data and expressing the data in such a way as to highlight their similarities and differences. Since in high dimension data it is hard to find patterns, where the luxury of graphical representation is not available PCA is a powerful tool for analyzing data. Once patterns have been extracted from the data, and one needs to compress the data (i.e. by reducing the number of dimensions) without much loss of information, PCA is a good choice for it. In terms of information theory the idea of using PCA is to extract the relevant information in a face image, encode it as efficiently as possible and compare test face encoding with a database of similarly encoded models. A simple approach to extract the information contained in an image of face is to somehow capture the variations in a collection of face images independent of judgment of features and use this information to encode and compare individual faces [93]. In mathematical terms, the purpose of using PCA is to find the principal components of distribution of faces, or the eigenvectors of the covariance matrix of the set of face images, treating each image as a point (vector) in a very high dimensional space. The eigenvectors are ordered, each one accounting for a different amount of 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. Each eigenvector appears as a ghostly face. Each individual face can be represented exactly in terms of a linear combination of the eigenfaces and 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. If an image has dimensions of 112 x 92 and we have a subspace of 10304 dimensions then this image will be a point in the space as shown in Figure 4.9. Here X_1, X_2, …….X_M show the dimension of face space.

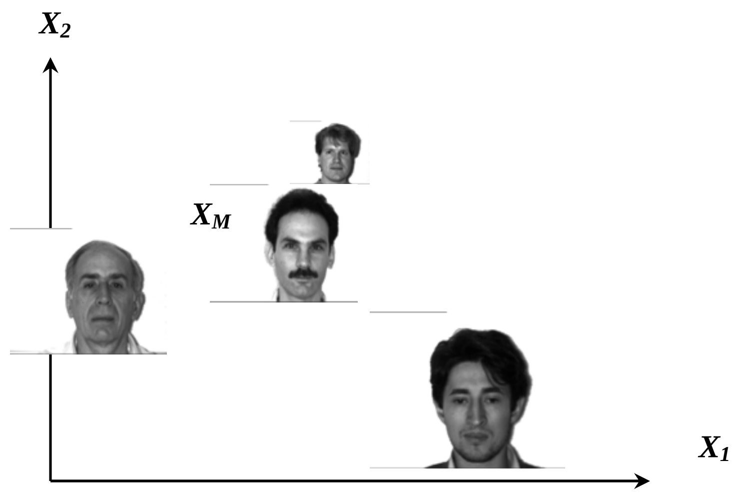


Figure 4.9. Four Different Face Projections in Face Space using PCA

The main use of PCA is to reduce the dimensionality of a data set while retaining as much information as possible. It computes a compact and optimal description of the data set. The first principal component is the combination of variables that explains the greatest amount of variation. The second principal component defines the next largest amount of variation and is independent of the first principal component. A mathematical model of PCA used for face recognition is shown in Figure 4.10 whose input is a set of training images $I=(i_1,i_2\ldots\ldots,i_n)$. The mean $\overline{m}$ of this training set is calculated, each image is centered by subtracting mean from it. This produces a dataset whose mean is zero. In next step, two dimensional variance called covariance of this dataset is calculated. As covariance matrix is a square matrix, its eigen values and eigen vectors are calculated which provide the information about patterns in the data. These eigen-values are ordered from highest to lowest and similarly the corresponding eigenvectors which provides data components in order of significance. This arrangement of data allows to decide on ignoring the data of less significance.

In this way, we do lose some information, but if the eigenvalues are small, we don't lose much and the final dataset will have lesser dimensions than the original. Finally, this reduced dimension data is transposed so that eigenvectors are in row, with most significant eigenvector at the top and multiplied by the transpose of centered image. This new data matrix is projection of face image in eigenface space.

4.13 Implementation of PCA Based Face Recognition Model

To assess the contribution of face image resolution on FRR, a PCA based recognition model is proposed where face images with varying resolution are applied as input to system. The algorithm developed for this purpose is as below:-

Begin

1. Input all training images
2. Carry out image preprocessing
   - Call scale normalization module
   - Call geometric normalization module
3. Call Resolution module

4. Calculate eigenvalues and eigenvectors

    - Calculate mean of training images
    - Carry out image centering
    - Find out Covariance matrix of centered images
    - Obtain eigenvalues and eigenvectors of covariance matrix

5. Arrange the eigenvalues and corresponding eigenvectors in ascending order

6. Carry out dimension reduction through selection of highest eigenvalues and eigenvectors

7. Made face image projection in PCA subspace

8. Carry out image Recognition

    - Load test image
    - Repeat the steps 2 to 7
    - Obtain Euclidean distance of test projection with training images projection
    - Find out closest match

9. Display image with closest match

End

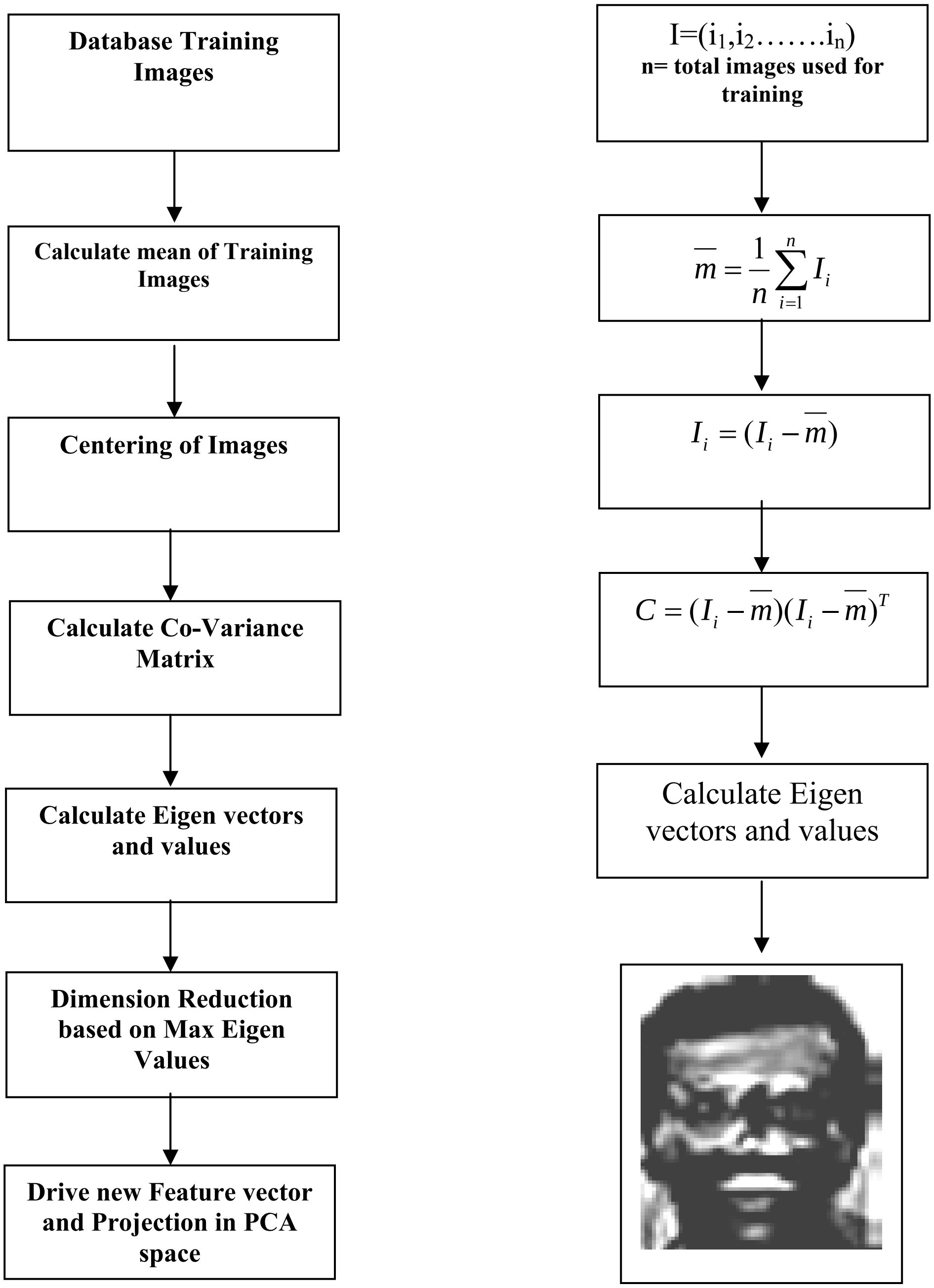


Figure.4.10.Illustration of Classic PCA

Five out of ten images of ORL database are used to learn the model. Scale and geometric normalizations are applied to the training images. Image resolution is varied, dimension reduction through PCA is carried out and face images are projected in PCA subspace as illustrated in Figure 4.10. Once the test image is presented to model in recognition process, it is preprocessed, its resolution is varied according to resolution of training images and its feature vector obtained through PCA is projected in the PCA subspace. Euclidian distance criterion is used to get the best match. Training and recognition process is shown in Figure 4.11.

4.14 Experiments and Results

Experiments have been conducted by varying the image resolution through image decimation as well as on actual resolution on which image have been acquired. In all these tests, five out of ten images are used for training purposes and complete database images have been used as test images. The number of eigenvectors (against highest eigenvalues) were retained according to number of image classes used in the model. Results have also been gathered by varying the number of classes used in the model as shown in Table 4.5.

Table.4.5. FRR of ORL Database against Varying Image Resolution

Resolution Down Scale Factor (G)	Image Resolution	Face Recognition Rate (%)	Improvement against normal PCA at normal resolution (%)
1	112x92	84	Nil
2	56x46	87	3
4	28x23	85	1
6	18x15	78	-6
8	14x11	73	-11

Results shown in Table 4.5 conclude that when image resolution is reduced, FRR is changed and attain its best value at a particular resolution. Beyond this value FRR becomes very poor. ORL database images are acquired at resolution of 112 x 92, Table 4.5 reflects that at this resolution FRR is 84%. As we decrease the resolution the FRR is improved and resolution of 56 x 46 (at value of G=2) gives best FRR of 87%. A further decrease in image resolution starts causing adverse effects on FRR. It proves that in ORL database a resolution of 56 x 46 retains the required image information for recognition whereas below or above this resolution either image loses the information required for recognition or presents unnecessary clutter of information giving wrong matching. Table 4.5 also represents the comparison of PCA based FRR obtained through varying image resolution with FRR gathered at normal resolution.

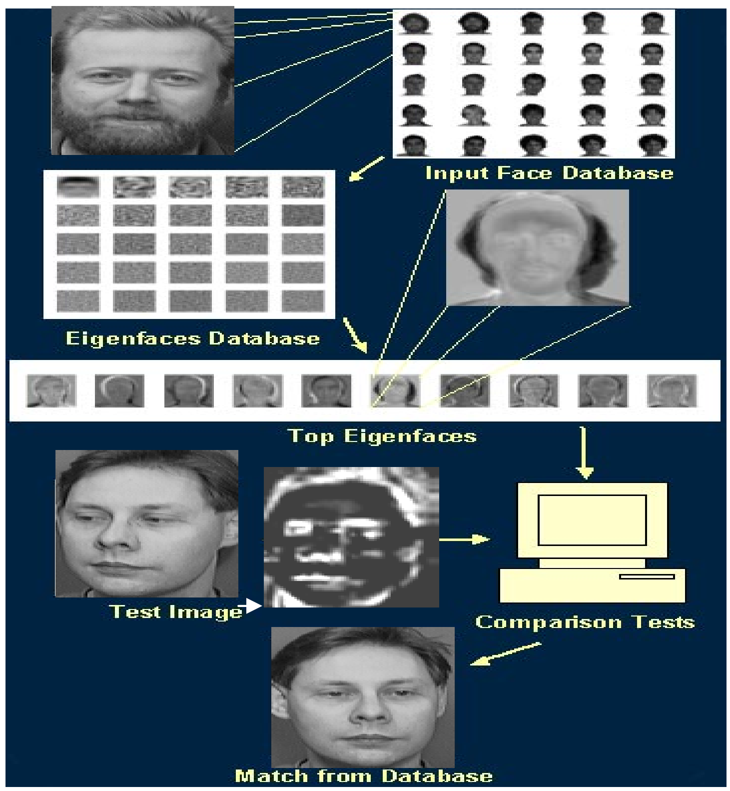


Figure.4.11 Face Recognition and Training Process using PCA

Afterwards, experiments were also conducted by keeping this reduced image resolution constant and varying number of classes used for experiments. A comparison of these results with normal PCA is carried out and is shown in Table 4.6 and Figure 4.12. This comparison demonstrates that as the numbers of classes are increased i.e. number of face representation in PCA subspace is increased, the FRR is reduced due to increase in the image clusters in the PCA domain and close placement of discriminating faces in PCA space.

Table.4.6. Comparison of FRR of ORL Database between normal and reduced image resolution

No of subjects used	FRR (%) with Normal Image resolution (112x92)	FRR(%) with Reduced Image resolution (56x46)	Improvement with column 2 (%)
10	97	97	Nil
15	97	98.5	1.5
20	92	94.5	2.5
25	91	94	3
30	89	93	4
35	86	89.75	3.75
40	84	87	3

The best FRR of 98.5% is obtained when only 15 subjects are used in training and recognition process. As the numbers of classes are increased from 15 the FRR is reduced because above 15 classes the projection of face images in PCA subspace becomes closer and closer and chances of mismatched are more as compared to less number of classes used.

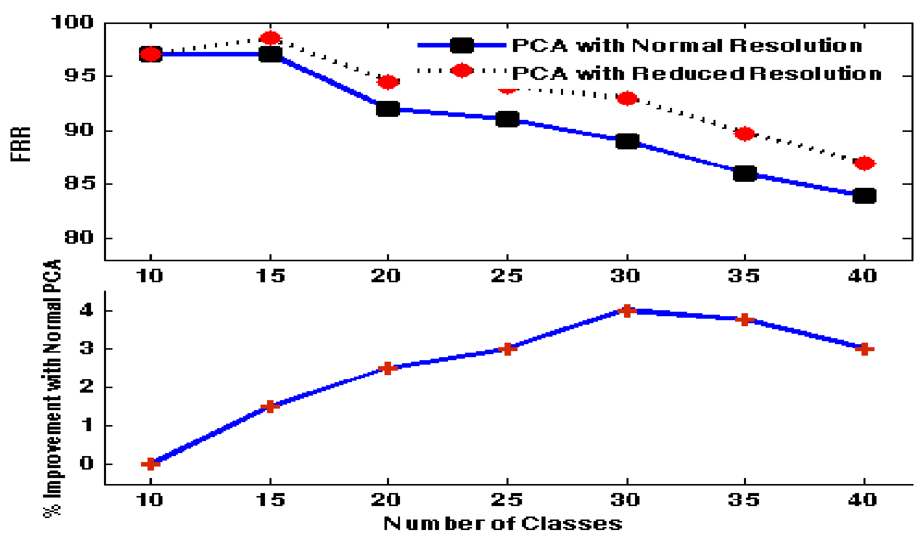


Figure.4.12. FRR with PCA for normal and reduced image resolution

4.15 Sub-Holistic PCA (SHPCA) Recognition Technique and Varying Image Resolution

The proposed technique is a simple offshoot of PCA where face image is split up into four parts as shown in Figure 4.13. Each of the four images formed are complete images and are a sub-set of the main image and called sub-images or sub-holons. Due to these sub-holons and mathematical nature of PCA, this technique is given name of sub-holistic PCA.

Figure 4.13. Image and the four sub-Images

This hybrid approach is based on two-step face recognition. The first recognition stage reduces the number of test subjects to five and the second recognition stage searches the best match among these five subjects. This concept is a common practice in daily life as well as in intelligent systems. Instead of finding a single match to the face, a number of matches are found and from among these candidate matches the final result is sought.

4.16 Implementation of SHPCA

In this hybrid approach of face recognition first step is to reduce the image resolution up to level where best FRR is achieved. Table 4.6 reflects that in case of normal PCA, the best FRR of ORL database is gathered at resolution of 56x46 so the resolution of all the training images is reduced to this level. This reduction in image resolution causes blur in the face images which is reduced by applying a high pass filter. Complete image is used to create one database and the four sub-images obtained from the single image are used to create the other four face spaces. The five face spaces can be differentiated as top left corner face space, top right corner face space, bottom left corner face space, bottom right corner face space and full image face space. As the name suggests the images formed by taking the top left corner portion of the images are used in the creation of the top left corner face space and so on as shown in Figures 4.14 and 4.15. Since the sub-face images formed are from within the larger image, they can be thought of as sub sets of the original image.

Training in SHPCA consists of two phases rather than a single phase as was the case in PCA. Each face space is created, with its respective data, similar to the creation of face space in PCA. Figure 4.14 describes the implementation of this proposed model of face recognition. The first step is to reduce the image resolution through image decimation as discussed earlier. Resultant image is divided into four sub images. The mean vectors are calculated by using expressions given in Equation 4.10 $\bar{\mathbf{E}}_{\mathbf{fs1}}$ is calculated for the top left corner face space. $\bar{\mathbf{E}}_{\mathbf{fs2}}$ is calculated for the top right corner face space. $\bar{\mathbf{E}}_{\mathbf{fs3}}$ is calculated for the bottom left corner face space. $\bar{\mathbf{E}}_{\mathbf{fs4}}$ is calculated for the bottom right corner face space. $\bar{\mathbf{E}}_{\mathbf{fs5}}$ is calculated for the full image face space.

$$\left\{\begin{matrix} \overline{E}_{fs_1} = \frac{1}{n}\sum_{i=1}^{n} E_{fs_1} \\ \overline{E}_{fs_2} = \frac{1}{n}\sum_{i=1}^{n} E_{fs_2} \\ \overline{E}_{fs_3} = \frac{1}{n}\sum_{i=1}^{n} E_{fs_3} \\ \overline{E}_{fs_4} = \frac{1}{n}\sum_{i=1}^{n} E_{fs_4} \\ \overline{E}_{fs_5} = \frac{1}{n}\sum_{i=1}^{n} E_{fs_5} \end{matrix}\right\} \tag{4.11}$$

An example of mean calculation of few sub-images is shown in Figure 4.15. In next step, each sub-image is translated by using Equation 4.12 so that mean of complete dataset is zero.

$$\begin{cases} E_{fs_1i} = E_{fs_1i} - \overline{E}_{fs_1} \\ E_{fs_2i} = E_{fs_2i} - \overline{E}_{fs_2} \\ E_{fs_3i} = E_{fs_3i} - \overline{E}_{fs_3} \\ E_{fs_4i} = E_{fs_4i} - \overline{E}_{fs_4} \\ E_{fs_5i} = E_{fs_5i} - \overline{E}_{fs_5} \end{cases} \tag{4.12}$$

The resultant covariance matrices, $\mathbf{C_{fs1}}$, $\mathbf{C_{fs2}}$, $\mathbf{C_{fs3}}$, $\mathbf{C_{fs4}}$ **&** $\mathbf{C_{fs5}}$, are used to calculate the eigenvectors $\boldsymbol{\lambda_{fs1i}}$, $\boldsymbol{\lambda_{fs2i}}$, $\boldsymbol{\lambda_{fs3i}}$, $\boldsymbol{\lambda_{fs4i}}$, $\boldsymbol{\lambda_{fs5i}}$ and eigenvalues $\boldsymbol{\Phi_{fs1i}}$, $\boldsymbol{\Phi_{fs2i}}$, $\boldsymbol{\Phi_{fs3i}}$, $\boldsymbol{\Phi_{fs4i}}$ & $\boldsymbol{\Phi f_{s5i}}$. The eigenvectors obtained are reordered in a descending eigenvalue order. The eigenvectors corresponding to the largest eigenvalues are retained because only they account for the most variance within the set of face images. Once the images have been projected into the five face spaces the system is considered as trained and it can be used for testing.

The test phase of the scheme can be divided into four phases. During the first phase, the image to be tested is passed through the same pre-processing procedure as was adopted for the images of the training phase. This results in a reduced size and sharpened image with four sub-images. The complete face image and the four sub-images are projected in their respective face spaces through PCA by pursuing the following steps:

The covariance matrices of each subspace are calculated by:

$$\begin{cases} C_{fs_1} = (E_{fs_1i} - \overline{E}_{fs_1})(E_{fs_1i} - \overline{E}_{fs_1})^T \\ C_{fs_2} = (E_{fs_2i} - \overline{E}_{fs_2})(E_{fs_2i} - \overline{E}_{fs_2})^T \\ C_{fs_3} = (E_{fs_3i} - \overline{E}_{fs_3})(E_{fs_3i} - \overline{E}_{fs_3})^T \\ C_{fs_4} = (E_{fs_4i} - \overline{E}_{fs_4})(E_{fs_4i} - \overline{E}_{fs_4})^T \\ C_{fs_5} = (E_{fs_5i} - \overline{E}_{fs_5})(E_{fs_5i} - \overline{E}_{fs_5})^T \end{cases} \tag{4.13}$$

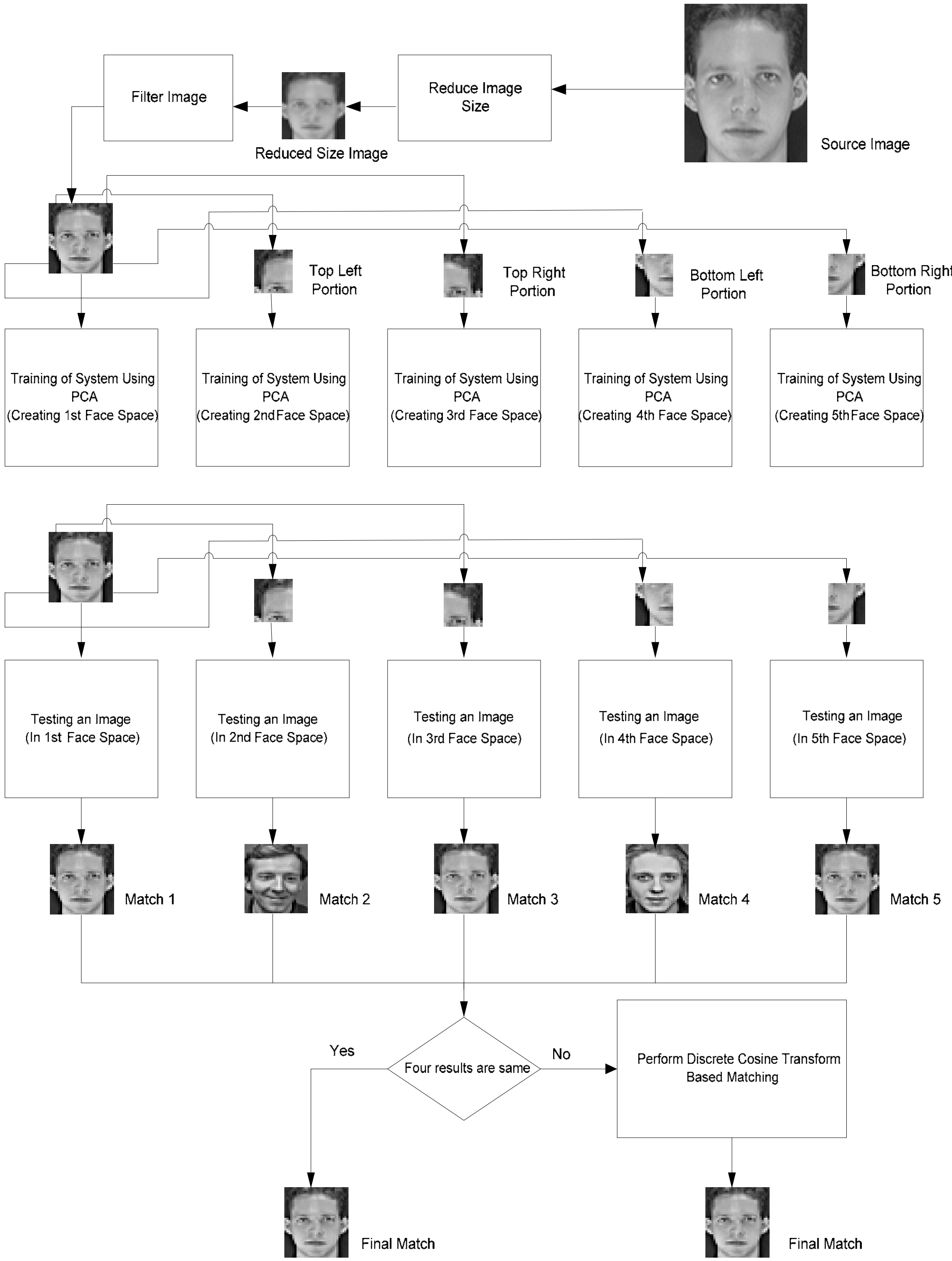
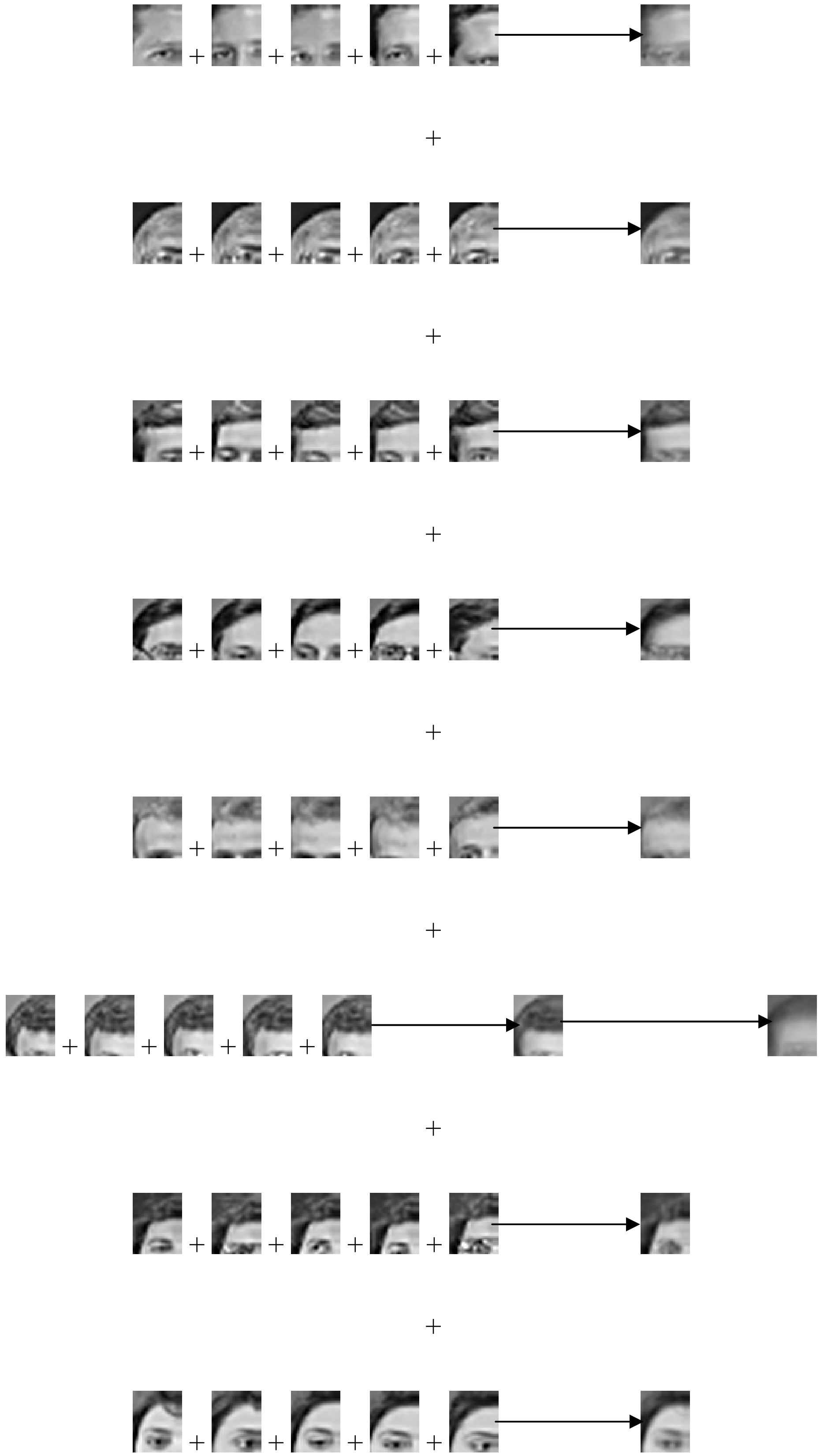


Figure.4.14. Representation of Proposed Face Recognition Model



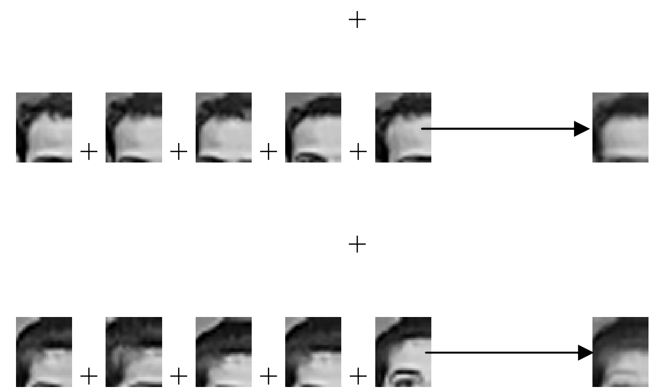


Figure. 4.15. Determination of Mean Face Image for the Top Left Corner Face Space

4.16.1 The mean face image is subtracted from the test image. This results in the translated test image. Mathematically:

$$\left\{\begin{aligned} T_{fs_1} &= T_{fs_1} - \overline{E}_{fs_1} \\ T_{fs_2} &= T_{fs_2} - \overline{E}_{fs_2} \\ T_{fs_3} &= T_{fs_3} - \overline{E}_{fs_3} \\ T_{fs_4} &= T_{fs_4} - \overline{E}_{fs_4} \\ T_{fs_5} &= T_{fs_5} - \overline{E}_{fs_5} \end{aligned}\right\} \tag{4.14}$$

4.16.2 The mean face image for the top left corner face space is subtracted from the top left corner of the test face image, the mean face of the top right corner face space is subtracted from the top right corner of the test face image and so as mentioned in Equation 4.14. Finally, the mean face of the full image face space is subtracted from the test face image. The resultant normalized or translated images are shown in Figure 4.16. The images on the left are the test images and the images on their right are the translated versions.

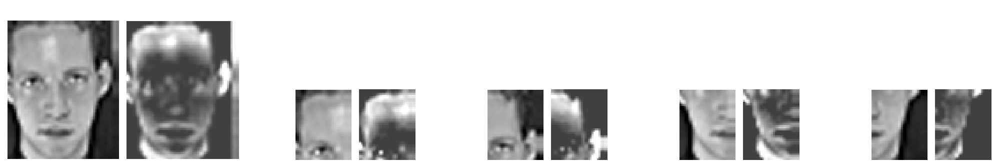


Figure 4.16. Test Image before and after translation

4.16.3 The translated images are then projected into their respective face space making use of the eigenvectors calculated during training phase. The minimum Euclidean distance is calculated between the projected test image and the training images already present in the face spaces. For each face space, the class/subject that returns the minimum difference is selected as the best match for that face space.

For the test image of Figure 4.17 the following matches were obtained:

- Top Left Corner Face Space Class / Subject 1
- Top Left Corner Face Space Class / Subject 1
- Top Left Corner Face Space Class / Subject 2
- Top Left Corner Face Space Class / Subject 8
- Top Left Corner Face Space Class / Subject 2

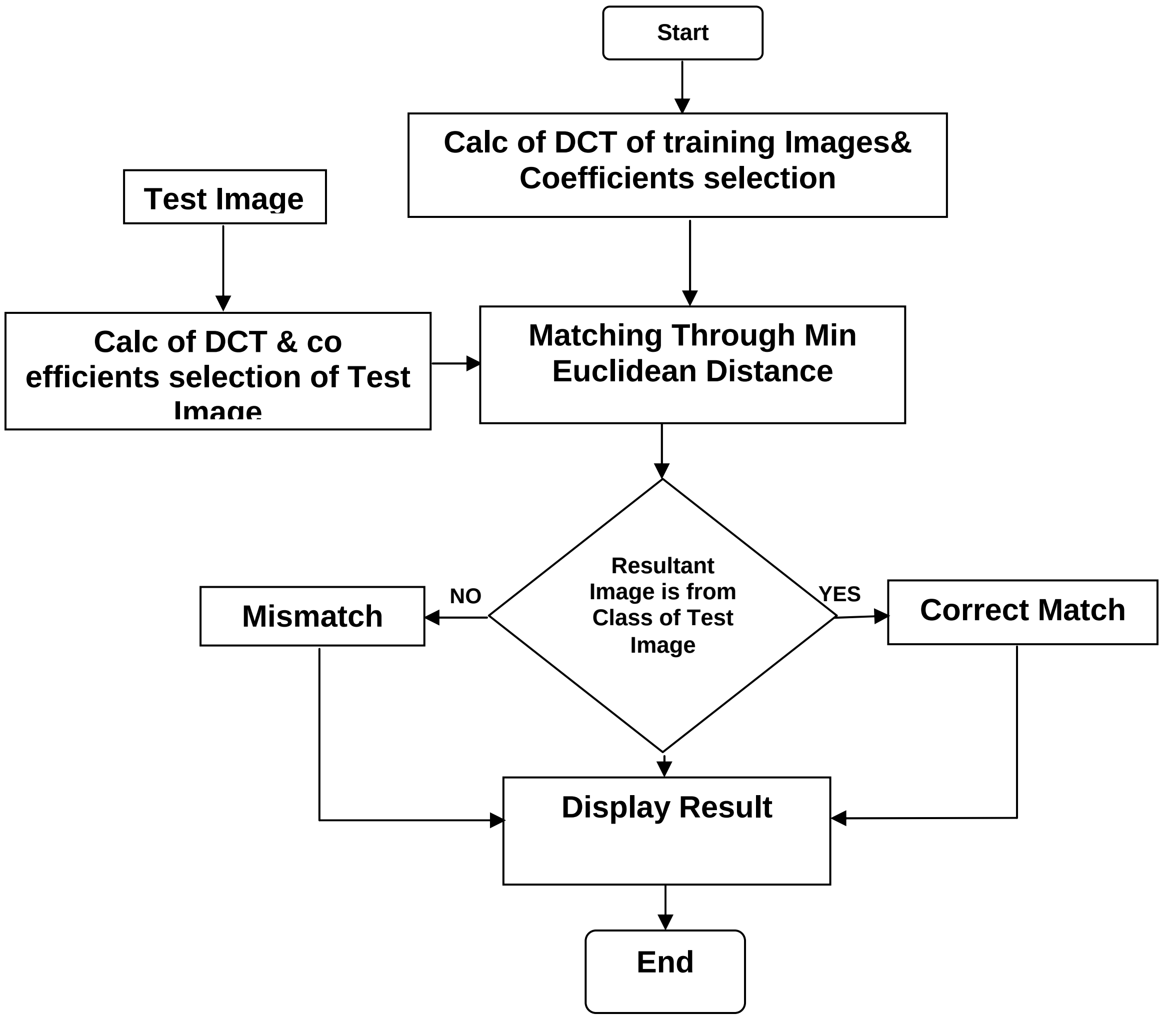
Thus, the five face spaces suggest that the test image could belong to the following three classes or subjects.

Test Image Subject 1 Subject 1 Subject 2 Subject 8 Subject 2

FaceSpace1 FaceSpace2 FaceSpace3 FaceSpace4 FaceSpace5

Figure 4.17 Test Image and the Results

Thus, the second phase results in five possible matches of the test image, one from each of the five face spaces. If four of the face spaces produce the same subject as the best match then the next two phases of the testing process are skipped and the result of the four face spaces is considered to be the best match. Otherwise, the five resultant subjects/classes are used as training set for the DCT based classification as described in flow chart 4.1.



Flow Chart No 4.1

In DCT based classification, (Flow Chart 1) the 2-dimensional DCT of the training set images is calculated and the coefficients are compared with the DCT coefficients of the test image. This is done by calculating the Euclidean Distance between the test image coefficients and the coefficients of the training set images. Minimum Euclidean distance is considered as the best result and the corresponding subject is considered to be the best match.

4.17 Experiments and Results

The proposed face recognition technique of SHPCA was tested and compared with PCA under two different set of conditions. In the first testing phase, the numbers of images of each class were fixed to five and the system was tested after training it with different number of subject classes. The number of classes were increased in steps of 5 from 10 to 40. The resultant comparison between standard PCA and SHPCA is shown in Table 4.7 and Figure 4.18.

Table 4.7. Comparison between standard PCA and Sub-Holistic PCA

No of Classes	FRR for Normal PCA	FRR for SHPCA
10	97 %	99 %
15	97 %	99 %
20	92 %	98 %
25	91 %	98 %
30	89 %	97 %
35	86 %	96 %
40	84 %	94 %

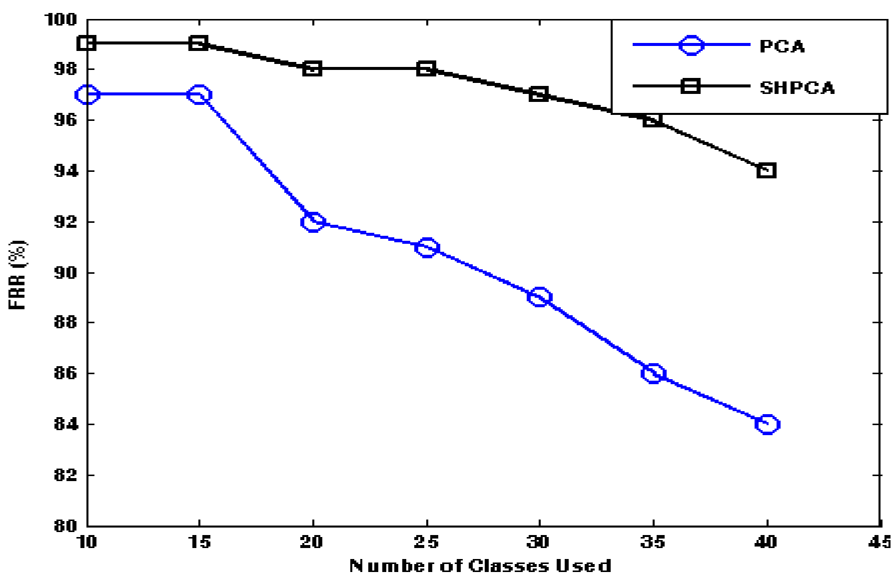


Figure 4.18. Comparison of PCA and SHPCA Recognition Rate for

5 training images of each class

In second testing phase, number of training images of each class is reduced from five to three. Proposed face recognition model was tested by increasing the number of classes used from 10 to 40 in a step of 5. Results are shown in Figure.4.19.

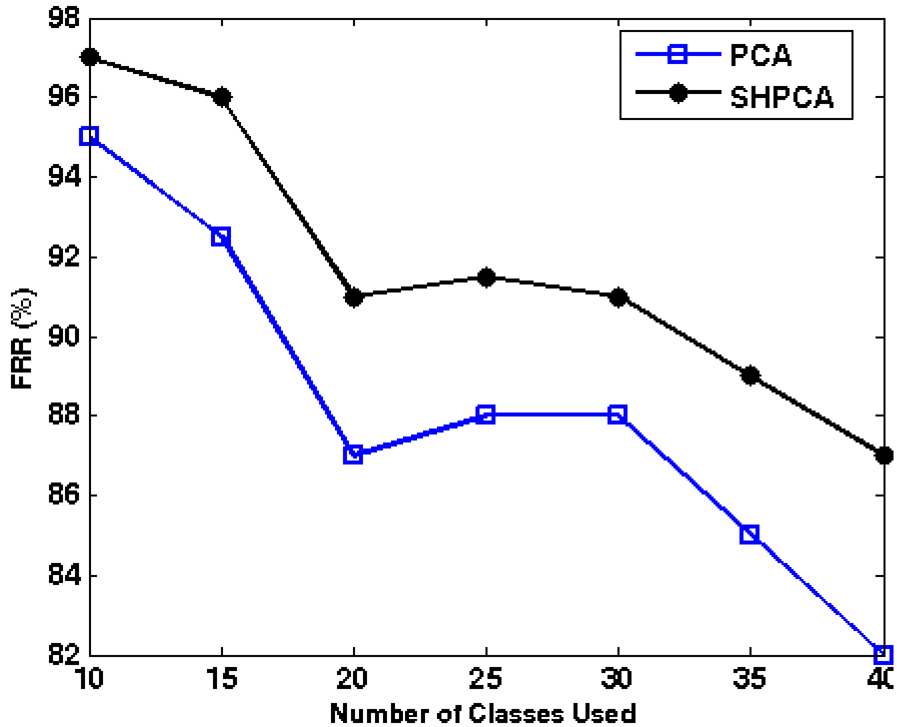


Figure 4.19. Comparison of PCA and SHPCA Recognition Rate for

3 training images of each class

4.18 Discussion of PCA and SHPCA Results

PCA is a set of variables that defines a projection which encapsulates the maximum amount of variation in a dataset and is orthogonal (and therefore uncorrelated) to the previous principal component of the same dataset. It is commonly used in micro array research as a cluster analysis tool. It is designed to capture the variance in a dataset in terms of principal components. In fact, one is trying to reduce the dimensionality of the data to summarize the most important (i.e. defining) parts while simultaneously filtering out noise. Same is also being used in this model for face recognition through reduction of dimension and projection of variances in face images.

Table 4.5 illustrates that image resolution has a significant effect on FRR. At a particular resolution, facial features within a class become promising to each other. Posture of certain image information changes with variation in image resolution and becomes more pledging with minimum difference within own class for recognition at a particular resolution. With a change in image resolution, the discriminative information possessed by face images are also changed whereas structure information remains same in all images at one particular resolution which gives best FRR. Table 4.5 describes that with the reduction in image resolution FRR also changes and attains its best FRR at a particular resolution (56x46 in case of ORL database). This phenomenon supports Shannon sampling theory which states that up to certain resolution a face image retains all its original information which depicts that if image is reconstructed from this resolution it does not lack any original information. Further reduction in image resolution beyond this value causes negative effect on FRR because below this particular image resolution requisite image information required for recognition is lost.

In Table 4.6, results were gathered by keeping the image resolution (56 x 46) against best FRR constant and number of image classes were varied. As the number of classes used are increased, FRR is reduced due to close representation of face images in PCA subspace which increases the chance of wrong matching and resulting reduction in FRR.

In mathematical model of PCA covariance matrix is the base of calculations of eigenvalues and eigenvectors. It can be defined as tendency of two datasets to vary together and a covariance matrix is merely collection of many co-variances:

$$C = (I_i - \overline{m})(I_i - \overline{m})^T \tag{4.15}$$

Once image resolution is reduced up to certain factor it reduces the high frequency clutter from the image not required for face recognition. Variance of this reduced data matrix provides better spread of variant data by removing undesired values from data matrix. Reduction in face data matrix should be such that if image is reconstructed from this resolution it does not lack any original information. The spread plot of one image with different resolutions is shown in Figure 4.20.

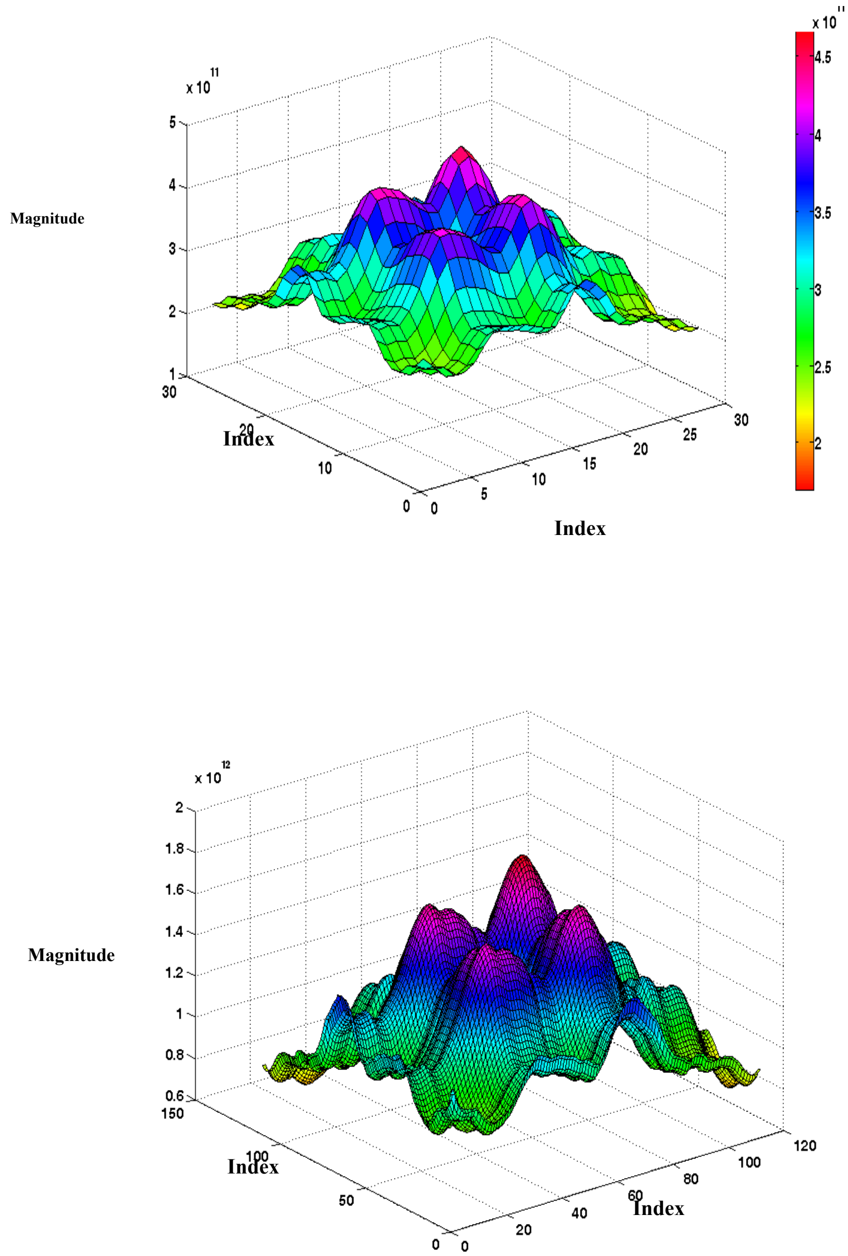


Figure. 4.20. Covariance spread with different image resolutions 56 x 46 (upper) and with 112 x 92 (lower)

Similarly, reduction in unnecessary data from covariance matrix correspondingly provides better eigenvalues choice as shown in Figure 4.21.

It also depicts that eigenvalue graph of reduced image resolution provides smooth elbow curve with better choice of few eigenvalues with maximum variance in images. It results in enhancing matching and recognition results. This particular image resolution level where reconstruction of image is perfect without loss of much facial information provides better eigenvector choice without any undesired clutter information. Moreover better choice of eigenvectors against narrow spread of eigenvalues provides improved FRR. Experiments on different resolution levels have been carried out and it is found that in ORL database two times reduction in resolution provide best FRR of 87%. This image resolution was also tested against change in number of classes used and results are shown in Figure 4.12 and Table 4.6.

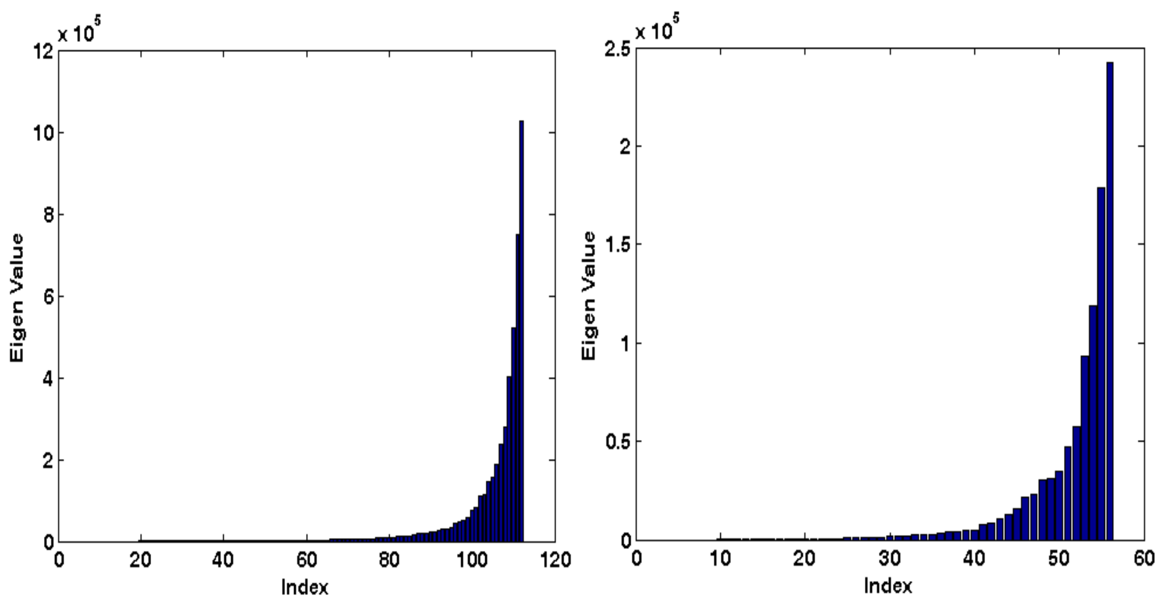
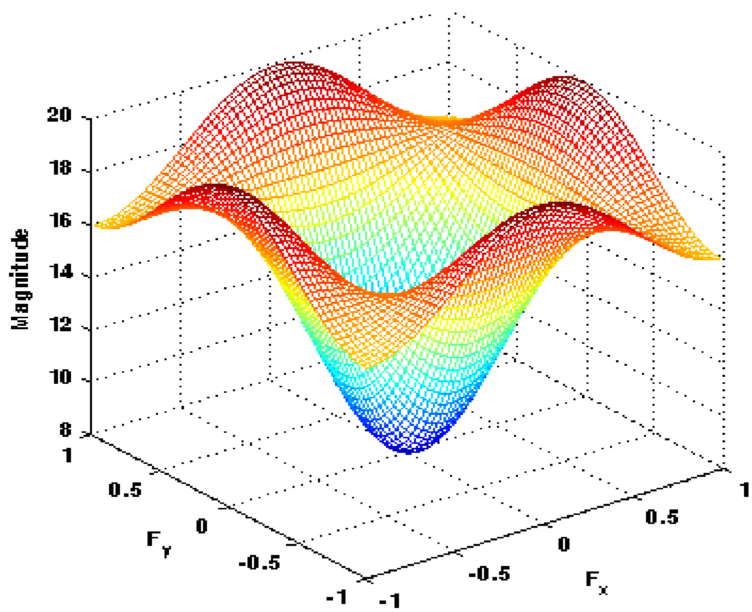


Figure. 4.21. Plot of Eigenvalues of with 112 x 92(right side) and with 56 x 46 (left side) Image Resolution

SHPCA is a new methodology used to improve the recognition rate as compared to standard PCA. First, image resolution was reduced to level where classic PCA achieved best FRR as discussed in previous classic PCA approach. Once image resolution through decimation was reduced it created blurring effect in the image which was removed through a sharpening filter shown in Figure 4.22.

Image was divided into four parts (Figure 4.15), a two step face recognition approach is used to improve the image classification. In first step, PCA was individually applied on these fours parts of image and on complete image as well and recognition was performed by using all these five parts. If decision is not yet made then DCT is applied to get best recognized image.



-1	-1	-1
-1	16	-1
-1	-1	-1

Figure 4.22. Sharpening Filter (right) and its Frequency Response (down)

Figure 4.23. Images before and after application of sharpening filter

The creation of the five face spaces in SHPCA results in increasing the training time to 50 seconds, compared to PCA's single face space training time of 30 seconds for all 40 subjects of the ORL dataset. Although two step testing approach has increased the average recognition time per image to 0.66 seconds compared to PCA's average recognition time per image of 0.02 seconds for the complete ORL dataset. However the use of the proposed technique increases the correct detection percentage to 94% for 40 subjects, compared to standard PCA's correct detection percentage of 84% for 40 subjects. However, time delay can be reduced by incorporating better hardware.

4.19 Comparison with other PCA Based Face Recognition Techniques

A comparison of proposed face recognition techniques with current existing PCA based recognition techniques is shown in Table 4.8.

Table 4.8. Comparison of results of Proposed Face Recognition Model with Existing PCA based Recognition Techniques

Face Recognition Technique	**FRR**
A novel incremental principal component analysis and its application for Face recognition[94] (H.Zho August 2006)	92.50%
Weighted PCA space and its application in face recognition [95] (H.Y. Wang, August 2005)	88%
Face recognition using kernel principal component Analysis[96] (K.In.Kim 2002)	90.0%
Face recognition based on face –specific subspace[97] (S.Shan 2003)	88.4%
Two-dimensional Weighted PCA algorithm for Face Recognition[98] (V.D. M. Nhat,Jun, 2005)	95.06%
Proposed SHPCA face recognition model	94%

CHAPTER 5

TRANSFORM DOMAIN FACE RECOGNITION

5.1 Introduction

Dimension reduction and subspace analysis with improved results has been a major research issue in learning based image analysis, detection and recognition [99, 100, 101, 102, 103, 104]. One of these methodologies is frequency domain representation of face images which has two fold significance: (1) effective characterization of a pattern of interest, or effective classification of different patterns, and (2) dimension reduction. This investigation is based on the frequency space representation of facial images, which relies on a global transformation, (i.e. every pixel in the image contributes to each value of its spectrum). The frequency spectrum is a plot of energy against spatial frequencies, where spatial frequencies relate to the spatial relations of intensities in the image.

All face images possess high information redundancy and correlation which results in inefficiency in recognition. Transformation of facial images in frequency domain can be used to reduce image information redundancy because only a subset of transformed coefficients are necessary to preserve most desired and important features required for recognition. Earlier studies [105, 106] concluded that information in low spatial frequency bands play a dominant role in face recognition as low-frequency components contribute to the global description, while the high-frequency components contribute to the finer details [107]. In this research work, Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) based face recognition methods have been developed where low frequency feature extraction through novel methods is employed to achieve better results. Curse of dimensionality has been addressed through image resolution reduction and by adopting better feature extraction methodologies. Later on contribution of face image resolution on recognition is evaluated by using images of varying resolution as input to the model. Image resolution variation is a phenomenon which alters the image information by dumping the high frequency components through averaging (low pass) filter and then discarding the every next number of pixels according to desired resolution. This image resolution variation is called image decimation which has a definite impact on recognition, same conjecture has been proved and supported in these face recognition models.

5.2 Linear Transforms

The term “transform” means to change form or appearance. In terms of signal processing, a transform is normally a tool that is used to convert the signals from time domain or spatial domain to the frequency domain. There are various instances when it is pertinent to have the signal in the time domain and on the other instances it is important to have the signal in the frequency domain. For most of the image processing purposes it is better to have the signal in the frequency domain. In others words, a transformation can be described as the process of mapping the correlated data to no-correlated data. Each pixel in an image is correlated with its neighbor pixels. The information represented by any pixel should be predicted by its neighbors because of the fact that they are all correlated.

5.3 Image Resolution Variation in Frequency Domain

In face images, resolution variation is one of the major contributors towards retaining suitable feature vector for recognition and dimension reduction, which enhances the computation speed and provides better image representation for further processing for recognition. Interpolation and decimation are the operations used to increase and reduce respectively, the sampled signals, usually by an integer factor. Enlargement of a sampled signal requires that new values not present in the signal be computed and inserted between the existing samples. The new value is estimated from a neighborhood of the samples of the original signal. Similarly, in decimation a new value is calculated from a neighborhood of samples and replaces these values in the minimized image. Image decimation is used to achieve images of different resolutions. Decimation is the elimination of data points from a data set. A decimation ratio of 4/5 means that 4 out of every 5 data points are deleted, or every fifth data point is saved.

Here in this model face image decimation is achieved by convolving the face image with an averaging filter and then decimator is used to discard pixels according to decimation down scale factor as shown in Figure 5.1(a). By varying the decimation downscale factor a Gaussian Pyramid of varying resolution of face images is obtained which is used in face recognition model to evaluate the effects of resolution on recognition in frequency domain.

Here *I(i,j)* as shown in Figure 5.1(b) is input image, *H* is convolution averaging mask and $C(n_1,n_2)$ is convolved image without zero padding. *F(m,n)* is the output decimated image with reduced image resolution shown in Figure 5.1(e).

$$F(m,n)=C[\,n_1G,n_2G] \tag{5.1}$$

where

G is decimation factor and

$$0 \le m \le (n_1/G) \quad , \quad 0 \le n \le (n_2/G) \tag{5.2}$$

The resulting image is a reduced size mirror of the original image faithful in tonality to the original but smaller in size.

5.4 Facial Features Representation through Discrete Fourier Transform (DFT)

The Fourier transform plays a critical role in a broad range of image processing applications including enhancement, analysis, restoration and recognition. DFT is used to decompose an image into its sine and cosine components. The output of the transformation represents the image in the Fourier or frequency domain, while the input image is the spatial domain equivalent. In the Fourier domain image, each point represents a particular frequency contained in the spatial domain image and it produces a complex number valued output image which can be displayed with two images, either with the real and imaginary part or with magnitude and phase. In image processing, often only the magnitude of the Fourier Transform is displayed, as it contains most of the information of the geometric structure of the spatial domain image. However, if we want to re-transform the Fourier image into the correct spatial domain after some processing in the frequency domain, we must make sure to preserve both magnitude and phase of the Fourier image.

DFT of a square image $g(m,n)$ with size m x n is:

$$G(u,v) = F\{g(m,n)\} \tag{5.3}$$

$$G(u,v) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} g(m,n) e^{-j2\pi\left(\frac{mu}{M}+\frac{nv}{N}\right)} \tag{5.4}$$

where Inverse DFT is:-

$$g(m,n) = F^{-1}\{G(u,v)\} \tag{5.5}$$

$$g(m,n) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} G(u,v) e^{j2\pi\left(\frac{mu}{M}+\frac{nv}{N}\right)} \tag{5.6}$$

where *g(m,n)* is the image in the spatial domain and the exponential term is the basis function corresponding to each point *G(u,v)* in the Fourier space. The equation can be interpreted as: the value of each point *G(u,v)* is obtained by multiplying the spatial image with the corresponding basis function and summing the result.

In general the Fourier transform is a complex function with a real and an imaginary part:

$$G(u,v) = R(u,v) + I(u,v) \tag{5.7}$$

The frequency, power and phase spectrum are given as:

$$\left\{\begin{aligned} & P(u,v) = R^2_{u,v} + I^2_{u,v} \\ & \theta = \tan^{-1}\frac{I_{u,v}}{R_{u,v}} \end{aligned}\right\} \tag{5.8}$$

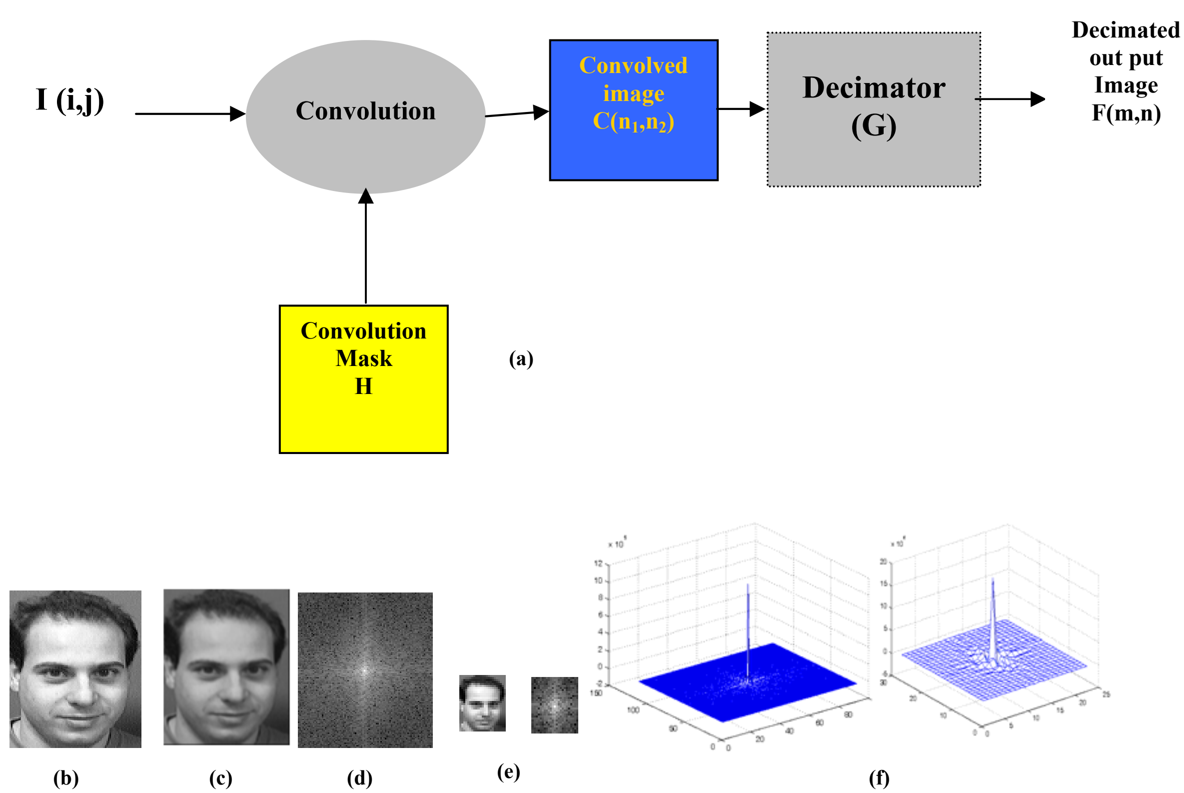


Figure 5.1. (a)Image Decimation Process, (b) Original Image, (c) Convolved image with Low pass Filter, (d) DFT spectrum of (c), (e)Result of Decimator (image and DFT spectrum) and (f)Plot of (d) and (e) spectrums

The DFT coefficients produced by the 2D DFT equation are arranged in a manner as shown in Figure 5.2.

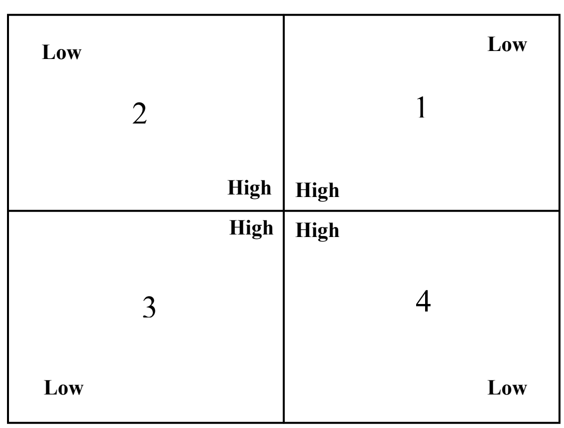


Figure 5.2. Frequency Contents Distribution of Fourier Transformed Image

It is considered much more intuitive to have low frequency contents in the center of the image spectrum and high frequency content on the outer sides of the image spectrum. Due to the periodicity of the contents and the fact that one could have done DFT over any period of the image and chosen to modify the frequency domain contents representation by interchanging the 1st and 3rd quadrants and 2nd and 4th quadrants. The basis functions sine and cosine waves are now with increasing frequencies, *i.e.* $G(0,0)$ represents the DC-component of the image which corresponds to the average brightness and $G(M-1,N-1)$ represents the highest frequency. In face recognition, low frequencies have the most important information which decreases as we move away from center of Fourier spectra of face image but contributes maximum in recognition. Whereas the higher frequencies which contribute towards finer details are less significant for recognition. Real part of low frequency components which contain maximum information as compared to imaginary part are considered for recognition purposes. The proposed DFT based recognition model is shown in Figure 5.3.

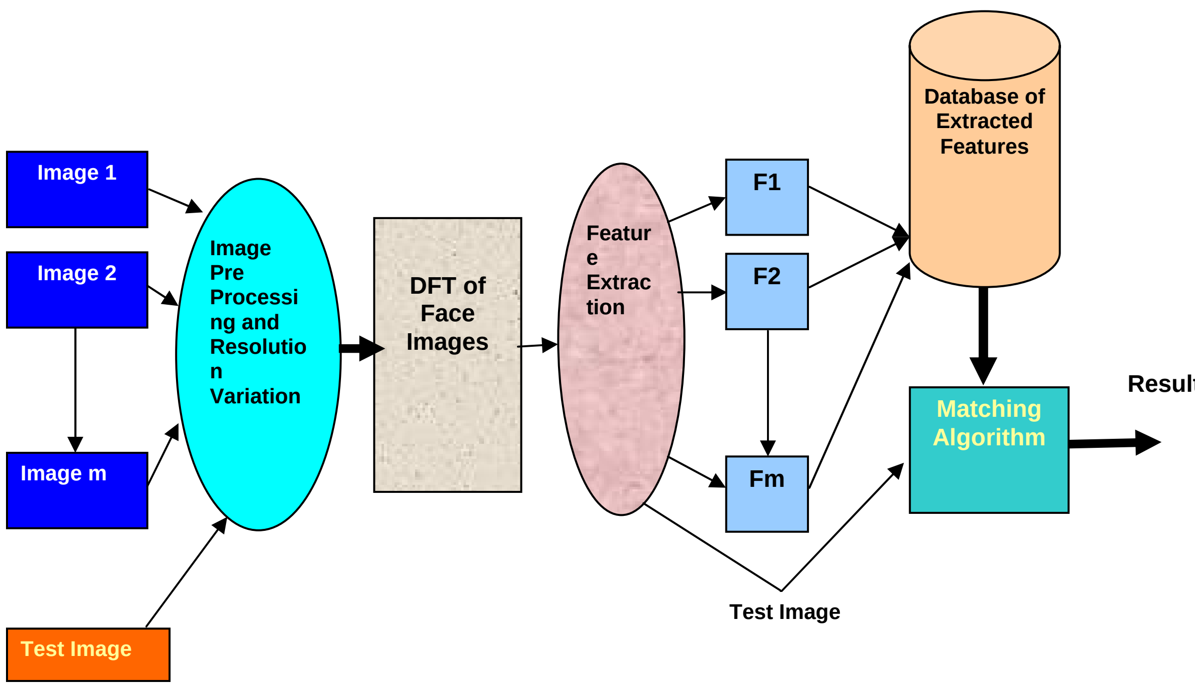


Figure. 5.3. DFT Based Face Recognition Model

5.5 Low Frequency Co efficient Selection Methods

Coefficient selection and dimension reduction are two main parameters in any pattern recognition technique. Feature selection is the choice of features which contribute maximum to recognition and dimension reduction is the minimum number of features required for better recognition rate and improved computation speed. Figure 5.1(d) shows the DFT spectrum of a face image which has components of all frequencies, low frequencies concentrate around the center but as we move away from spectrum center the magnitude gets smaller for higher frequencies. Hence, low frequencies contain more image information than the higher ones. The transformed image also reflects that there are two dominating directions in the Fourier image, one passing vertically and one horizontally through the center. These originate from the regular patterns in the background of the original image. In this DFT based face recognition research work, two distinct methods have been employed to extract best minimum feature vector from DFT spectrum of face image. These methods adequately represent the facial features required for better recognition. Moreover, theses methods improve the dimension reduction for better computation speed.

5.5.1 Square Selection Method

As low frequency components of image spectrum provide maximum desired facial information for recognition so a square of k x k dimension around the center of the spectra is selected and low frequency coefficients falling within this square are taken as feature vectors such that:

$$\left\{ \begin{matrix} k \in M, N \\ 0 < k < M, N \end{matrix} \right\} \tag{5.9}$$

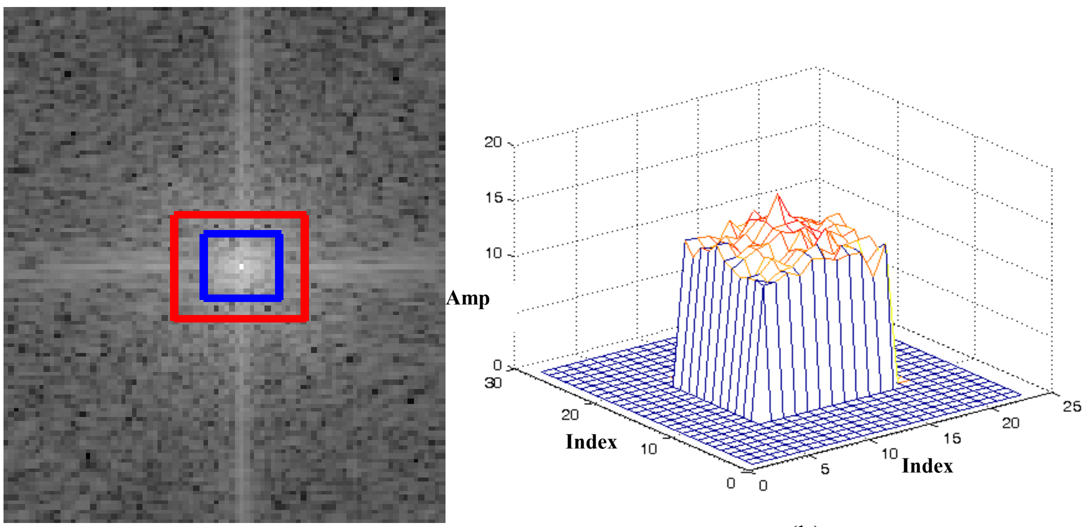
where

k is dimension of square

M, N is order of image spectrum matrix

Feature vector with different square dimensions is retained against each image in dataset.

Figure 5.4(a) represent the square selection of two different dimensions and correspondingly its plot is shown in Figure 5.4(b). Here coefficient selection is made by using the low frequency components falling within the square of specified length. Dimension reduction is achieved in this process as all coefficients except within square are discarded.



(a) (b)

Figure 5.4. (a) DFT spectrum with square selection around center (b) Plot of features extracted through square selection method

11.83	11.875	12.407	11.854	13.077	12.847	12.42	12.589	12.496	10.04	12.79
11.997	11.531	12.233	12.341	13.376	13.817	11.307	13.166	12.59	10.839	12.038
11.882	11.969	13.078	12.228	13.344	12.938	12.792	12.015	12.598	12.744	12.807
11.646	11.005	12.83	11.87	12.818	14.411	12.689	14.097	12.156	12.056	11.576
12.071	12.723	13.251	13.567	14.267	13.93	14.14	12.71	13.236	12.422	12.787
12.468	12.574	13.244	13.034	14.279	14.38	14.441	13.485	12.747	13.27	12.757
13.634	13.501	12.331	12.919	14.555	16.759	14.555	12.919	12.331	13.501	13.634
12.757	13.27	12.747	13.485	14.441	14.38	14.279	13.034	13.244	12.574	12.468
12.787	12.422	13.236	12.71	14.14	13.93	14.267	13.567	13.251	12.723	12.071
11.576	12.056	12.156	14.097	12.689	14.411	12.818	11.87	12.83	11.005	11.646
12.807	12.744	12.598	12.015	12.792	12.938	13.344	12.228	13.078	11.969	11.882

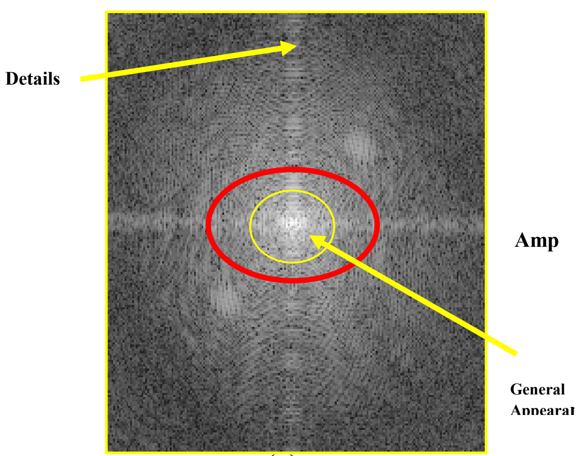
Figure 5.5. Real part of Pixel intensity value map with k = 5.

Figure 5.5 is the pixel intensity value map of ORL database image with square dimension of five. All the images in ORL database have been acquired at resolution of 112 x 92 i.e, total pixels of 10304, but by applying square selection method with square length of five only 121 pixel are used for recognition purposes, rest all are discarded which is a considerable dimension reduction.

5.5.2 Circular Selection Method

In this method, circles (Equation 5.10) with different radii (r) concentric with center of spectra are evaluated and feature vector of low frequency coefficients within circle is obtained.

$$\left.\begin{aligned} y &= r\sin\theta + y_c \\ x &= r\cos\theta + x_c \end{aligned}\right\} \tag{5.10}$$



(a)

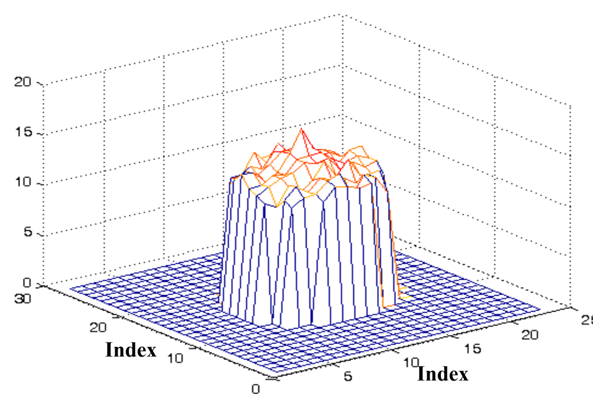

(b)

Figure 5.6. (a) DFT spectrum with circular selection around center (b) Plot of extracted features through circular selection method

where (x_c, y_c) are coordinates of center of circle, r is radius of circle and angle θ varies from 0 to 2π .

In this method coefficient selection is made through circle around the center instead of square. Figure 5.6(a) shows the circular selection of two different radiuses. Figure 5.6(b) depicts the plot of circular selected coefficients where as Figure 5.7 is the actual pixel intensity value map to represent the dimension reduction achieved through this method of feature extraction. These two methods of coefficient selection for recognition provide different dimensions of feature vectors with changing recognition rate. Varying image resolution is another factor which not only affects the recognition rate but also the computation speed which is another strong factor alongwith success rate.

			11.854	13.077	12.847	12.42	12.589			
	11.531	12.233	12.341	13.376	13.817	11.307	13.166	12.59	10.839	
	11.969	13.078	12.228	13.344	12.938	12.792	12.015	12.598	12.744	
11.646	11.005	12.83	11.87	12.818	14.411	12.689	14.097	12.156	12.056	11.576
12.071	12.723	13.251	13.567	14.267	13.93	14.14	12.71	13.236	12.422	12.787
12.468	12.574	13.244	13.034	14.279	14.38	14.441	13.485	12.747	13.27	12.757
13.634	13.501	12.331	12.919	14.555	16.759	14.555	12.919	12.331	13.501	13.634
12.757	13.27	12.747	13.485	14.441	14.38	14.279	13.034	13.244	12.574	12.468
	12.422	13.236	12.71	14.14	13.93	14.267	13.567	13.251	12.723	
	12.056	12.156	14.097	12.689	14.411	12.818	11.87	12.83	11.005	
			12.015	12.792	12.938	13.344	12.228			

Figure 5.7. Real part of Pixel intensity value map with r =5

5.6 DFT Based Face Recognition Process

Scale and geometric normalization of face images is carried out in pre-processing phase of this recognition model. A Gaussian pyramid of different resolutions is achieved through image decimation. Later on, DFT is applied on the images, low frequencies lie on the four corners of the DFT spectrum. To apply the desired coefficient selection methods with more precision, low frequency coefficients are shifted from spectrum corners to center of spectrum as shown in Figure 5.8.

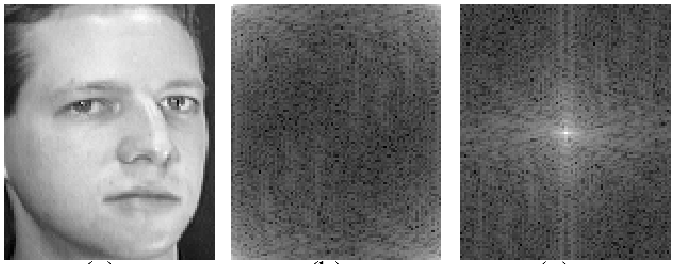


Figure.5.8. (a) Original Image, (b) DFT spectrum low frequency components at four corners, (c) DFT spectrum with low frequency components around the center

As imaginary part of DFT spectrum carries very little information required for recognition it is discarded and only real part is used. Now low frequency coefficients fall around the center of spectrum, feature selection through circular and rectangle method is made. This process reduces the dimension of the face while retaining the most important facial information required for recognition. A sub-space comprising of these feature vectors is retained as reference against each training face image.

Once a test image ***P*** is presented to the model, scale and geometric normalization is carried out and then feature vector through feature extraction is obtained. In matching process a dissimilarity space $D(T_i,P)$ of test image with training images is obtained by using simple Euclidean distance.

$$E = D\left(T_i, P\right) \tag{5.11}$$

This dissimilarity space matrix is converted to a vector with min difference is taken as matched image:-

$$R = arg\ min\left[E\right] \tag{5.12}$$

Where

P is test image

T is training image

i is 1 to total subjects used in training of model and

arg min is the recognized image.

Here the *arg min* which is accumulated minimum Euclidean distance provides the best match.

5.7 Implementation

An algorithm developed for DFT based recognition model is:

Begin

1. Input training images
2. Carry out image preprocessing
   - Call scale normalization module
   - Call geometric normalization module
3. Call Resolution module
4. Calculate DFT of the image
   - Shift the low frequency components around the center
   - Choose only real part of spectrum
   - Take log of spectrum

5. Extract feature vector using square and circular feature selection methods

6. Carry out dimension reduction through selection of low frequency coefficients

7. Convert extracted feature vector in single column

8. Carry out image Recognition

   - Load test image

   - Repeat the steps 2 to 7
   - Obtain Euclidean distance of test image feature vector with feature vector of training images
   - Find out closest match

9. Display image with closest match

End

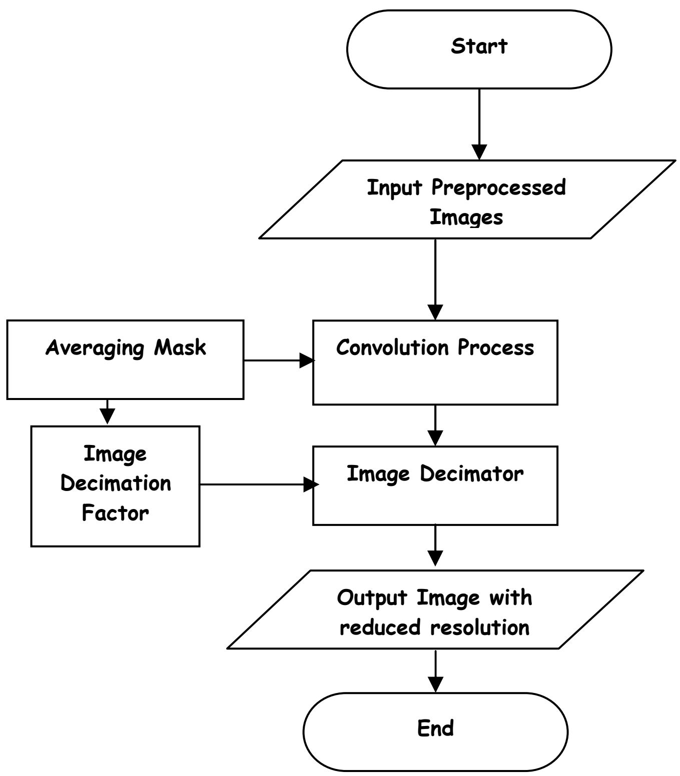
To implement this algorithm following different modules and data structures have been developed:

5.7.1 Preprocessing Module

In this module color images are converted into grayscale images, scale and geometrical normalization is carried out as discussed in preceding face recognition models.

5.7.2 Resolution variation Module

In this module image resolution is varied according to image decimation factor. Preprocessed images are loaded as input to this module as described in flow chart shown below. A 2D averaging mask is defined whose order is fed to decimator as image decimation factor. The input image is convolved with the averaging filter. Convolved blur image is provided to decimator. Function of decimator is to decimate the image by discarding the every next pixel according to decimation factor, resulting image is a reduced size and resolution image. The decimation factor is varied to obtain the images of varied resolutions but kept constant during one complete cycle of training and recognition process. Once the results on one particular image resolution are obtained then decimation down scale factor is changed to determine next set of results.



Flow chart of Image Resolution Variation Process

5.7.3 Feature Selection Module

In this module two different methodologies of feature extraction have been implemented.

5.7.3.1 Square Feature Selection Procedure

Images with a specific image resolution are loaded as input to in this procedure. Dimension of square is defined as second input to module. A square with a specific length is marked around the center of DFT spectrum of face image. Coefficients falling within this square are calculated and retained as a feature vector.

$$S_{(k\,x\,k)} \in I_{(M\,x\,N)} \tag{5.13}$$

where
S is feature vector
k is dimension of square
I is image with reduced resolution

Feature vector $S_{(k\ x\ k)}$ is converted to single column vector with dimension of $S_{(k^2\ x\ 1)}$. This single column feature vector against each image used in learning process of model is stored as reference. Five images of each class are used in learning phase of the model and a resultant matrix $R_{(k^2\ x\ T)}$ is obtained where T is total images used for training of the module. Same procedure is repeated with next dimension of square.

5.7.3.2 Circular Feature Selection Procedure

In this procedure of feature extraction, face images with specific resolution are loaded as input to module and a variable of radius of circle is defined as second input to procedure. The circle of specific radius around the center of transformed spectrum is calculated and low frequency coefficients falling within this circle are extracted as feature vector.

$$(C = 2\pi r^2) \in I_{M\,x\,N} \tag{5.14}$$

where
C *is feature vector*
r *is radius of circle*
I *is image with reduced resolution*

This feature vector against each image used in learning process of model is retained as reference. A matrix of dimension of $C_{(C \times T)}$ containing the extracted features of all training images is obtained whose each column represents the feature vector of a single training image. Process is repeated with variation of radius dimension but the radius of the vector is kept constant throughout a set of experiments.

5.7.4 Recognition Module

In recognition module, a preprocessed test image is taken as input. Its resolution through resolution module is changed according to resolution of training images. Feature vector (of same dimension as of feature vector obtained in learning process of the model) is extracted through feature extraction module. Euclidean distance as discussed in preceding face recognition model is used as image classification criterion.

5.8 Experiments and Results

Databases (ORL,YALE, and FERET) as discussed in chapter 4 have been used to evaluate the changes in FRR due to variation in image resolution and feature vector length. Results have been gathered in four different scenarios by using both methods of feature extraction.

5.8.1 Experiment Set 1

This set of experiments is based on square feature extraction method. Recognition results have been obtained by varying the feature vector length by changing the square dimensions. The results acquired on YALE database images (image resolution 56x46) against varying square dimension are shown in Table 5.1.

Table.5.1. FRR of Yale Database against varying Square Dimension

Square Dimension (k)	Training Time of database (Seconds)	Recognition Time Per image(Seconds)	Number of failure	FRR (%)
2	3.23	0.0430	10	93.9
3	3.37	0.0432	5	96.9
4	3.42	0.0437	2	98.7
5	3.45	0.0439	2	98.7
6	3.50	0.0441	1	99.3
7	3.56	0.0444	2	98.7
8	3.6	0.0455	4	97.5

165 images of Yale database provide best results of more than 99.3% at square dimension of 6. Dimension length "k" means k pixels on right and k pixels on left of center pixel of spectrum which means total length of feature vector will be:

$$Feature\ vector\ length = L^2$$
$$where$$
$$L = (k + k + 1)$$

Let k=5 then total number of pixels of feature vector as shown in Figure 5.5 are

$$Feature\ vector\ length = (5+5+1)^2 = 121$$
$$where\ \ k = 5$$

Table 5.1 reflects that as dimension of feature vector increases from 2, FRR improves and achieve best results at feature vector length of 6. Beyond this level, recognition rate suffers because increase in square dimension produces information redundancy in feature vector. This illustration shows that a specific number of feature coefficients are sufficient to represent the face image for recognition purposes.

Similarly results have been obtained on FERET and ORL database at image resolution of 128x193 and 56x 46 and are reflected in Table 5.2. Results illustrated in Table 5.2 and comparison given in Figure 5.9 reflect that each database provides best FRR on different square dimensions. FERET database provides 97.8% FRR at square dimension of 4, ORL database gives maximum FRR of 96.5% at square dimension of 5 where as YALE database achieved optimum FRR of 99.3% at feature vector dimension of 6 (Table 5.1).

Moreover, second column in Table 5.2 indicates the total number of pixels used for recognition against a particular length of k. It reflects the dimension reduction achieved through this feature extraction method.

Table.5.2. FRR of ORL and FERET Databases against varying Square Dimension

Square Dimension	**Total Number of Pixels**	**Training Time Seconds**		**Recognition Time Seconds**		**ORL Database**		**FERET Database**	
		ORL Database	**FERET Database**	**ORL Database**	**FERET Database**	**No. of Failure**	**FRR (%)**	**No. of Failure**	**FRR (%)**
2	25	18.2	38.49	41.89	125.93	22	94.5	39	96.1
3	49	18.64	39.21	43.80	129.07	21	94.75	30	97
4	81	19.01	39.9	44.2	132.9	16	96	22	97.8
5	121	19.60	40.2	44.5	137.6	14	96.5	23	97.7
6	169	19.83	40.8	45.1	142.87	16	96	25	97.5
7	225	20.02	41.2	45.5	146.1	20	95	26	97.4
8	289	22.01	42.1	46.1	152.5	23	94.25	29	96.9

This difference of square dimensions and best FRR registers that each database has different length of feature vector to represent effectively the facial information required for recognition.

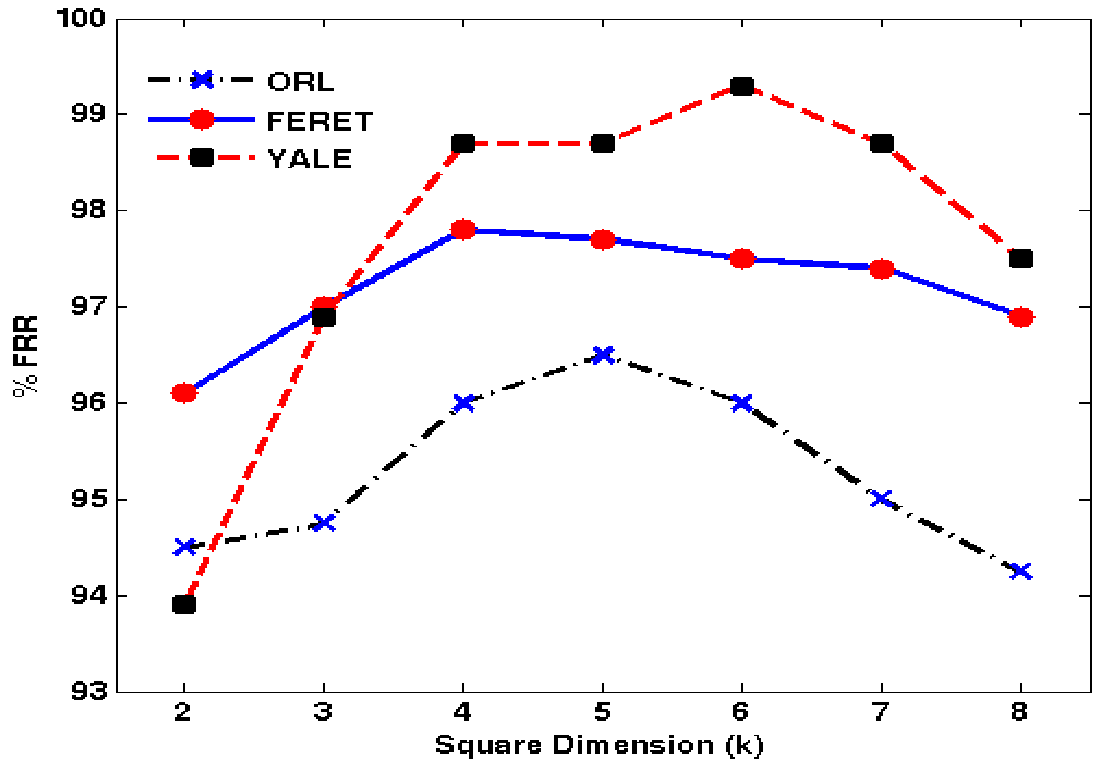


Figure.5.9. FRR of ORL, YALE and FERET databases against varying Square Dimension

5.8.2 Experiment Set 2

In this set of experiments results have been gathered by keeping the feature vector dimension constant against best FRR and varying the image resolution. Results of YALE database tabulated in Table 5.3 reflect that image resolution variation also affects the FRR. Yale database provides best result of 99.3% at resolution of 56 x 46.

Table 5.3. FRR of Yale Database with constant Square Dimension and varying image resolution

Image resolution	**Training Time (Seconds)**	**Recognition Time (Seconds)**	**No of Failure**	**FRR(%)**
112x92	10.04	15.21	3	98.1
56x46	3.13	7.21	1	99.3
28x23	2.26	6.35	4	97.5
18x15	1.2	5.92	5	96.9
14x11	1	4.7	6	96.3

Table 5.4. FRR of ORL and FERET Database with constant Square Dimension and varying image resolution

Image Resolution		**Training Time Seconds**		**Recognition Time Seconds**		**ORL Database**		**FERET Database**	
ORL Database	**FERET Database**	**ORL Database**	**FERET Database**	**ORL Database**	**FERET Database**	**No. of Failure**	**FRR (%)**	**No. of Failure**	**FRR (%)**
112x92	**256x384**	**37.3**	**94.86**	**51.01**	**232.34**	**22**	**94.5**	**31**	**96.9**
56x46	**128x192**	**19.5**	**29.34**	**44.5**	**156.32**	**14**	**96.5**	**22**	**97.8**
37x30	**85x128**	**19.0**	**17.21**	**42.0**	**142.2**	**17**	**95.75**	**24**	**97.6**
28x23	**64x96**	**18.1**	**13.25**	**41.6**	**138.9**	**18**	**95.5**	**30**	**97**
22x18	**51x76**	**14.6**	**15.2**	**38.2**	**123.7**	**21**	**94.75**	**33**	**96.7**
18x15	**42x64**	**13.2**	**14.0**	**37.01**	**112.3**	**23**	**94.25**	**41**	**95.9**

Results of FERET and ORL databases have been tabulated in Table 5.4 which indicates that ORL provides best FRR of 96.5 at decimation down scale factor 2 (image resolution of 56 x 46) and FERET database achieved optimum FRR at decimation down scale factor of 2 (image resolution of 128x 193). At the same time it is evident from results that as the image resolution is reduced by increasing decimation down scale factor it decreases the image dimension which ultimately reduce the training and recognition time of the model. This reduction in model processing time enhances the model computation speed.

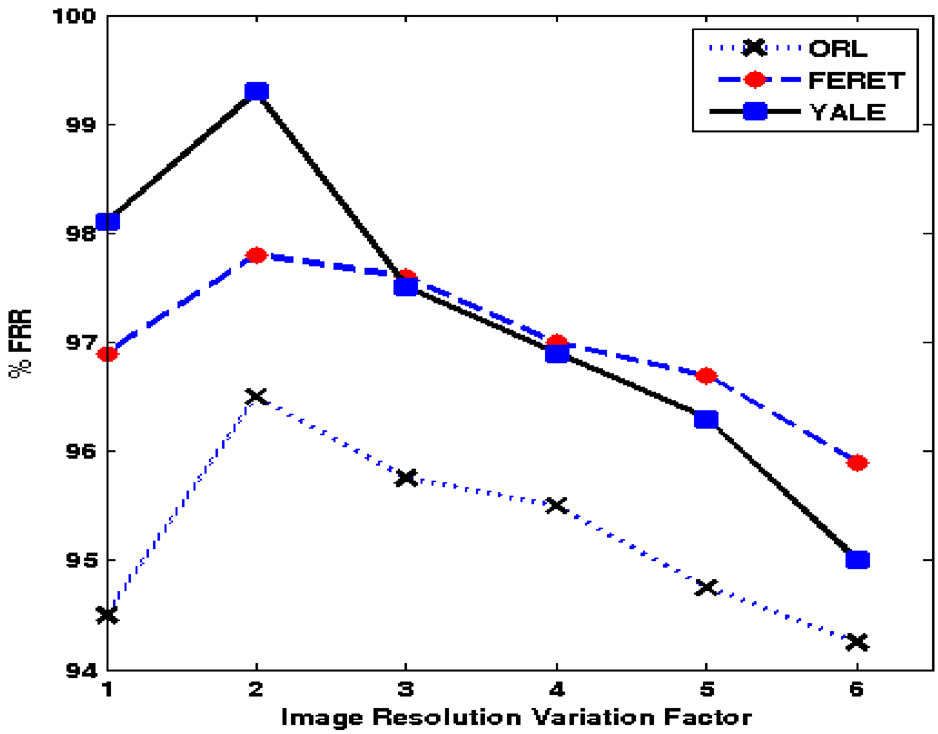


Figure.5.10. FRR with constant values of Square Dimension and varying Decimation Factor

5.8.3 Experiment Set 3

In experiment set 3, circle feature extraction method has been used to choose the best feature vector for recognition. Image resolution was kept constant throughout and dimensions of circle were changed to assess its effects on success rate. Table 5.5 and 5.6 show the results of YALE, ORL and FERET databases. It is evident from the results that as the radius of circle is increased, data in feature vector is augmented with required coefficients for recognition. As the dimension of feature vector is increased beyond a certain level the feature vector is polluted with redundant data, decreases the FRR resultantly.

Table 5.5. Yale Database Results with varying Circle Dimension

Circle Dimension (r)	Training Time of database (Seconds)	Recognition Time per Image (Seconds)	No. of Failure	FRR(%)
2	4.0	9.3	10	93.9
3	4.8	9.4	4	97.5
4	5.0	9.5	2	98.7
5	5.1	9.65	1	99.3
6	5.4	9.82	0	100
7	5.8	10.88	1	99.3
8	5.9	10.9	3	98.1

Table 5.6. ORL and FERET Database Results with varying Circle Dimension

Circle Dimension	**Training Time (Seconds)**		**Recognition Time (Seconds)**		**ORL Database**		**FERET Database**	
	ORL Database	**FERET Database**	**ORL Database**	**FERET Database**	**No. of Failure**	**FRR (%)**	**No. of Failure**	**FRR (%)**
2	**38..49**	**27.82**	**65.79**	**161.89**	**24**	**94.00**	**62**	**93.8**
3	**39.21**	**27.67**	**66.26**	**164.80**	**15**	**96.25**	**43**	**95.7**
4	**39.9**	**28.25**	**66.39**	**166.85**	**15**	**96.25**	**33**	**96.7**
5	**40.2**	**28.92**	**66.59**	**167.82**	**12**	**97**	**23**	**97.7**
6	**40.7**	**30.20**	**67.22**	**170.4**	**11**	**97.25**	**36**	**96.4**
7	**41.4**	**31.03**	**68.12**	**171.5**	**20**	**95.**	**50**	**95**
8	**42.2**	**32.5**	**69.03**	**177.3**	**22**	**94.5**	**50**	**95**

In Figure 5.11 a comparison of FRR against varying circle dimension of these databases is shown. It indicates that Yale database attains 100% FRR at radius dimension of six. ORL database provides 97.25% at radius dimension of six whereas best FRR of 97.7% of FERET database has been observed at circle radius of five.

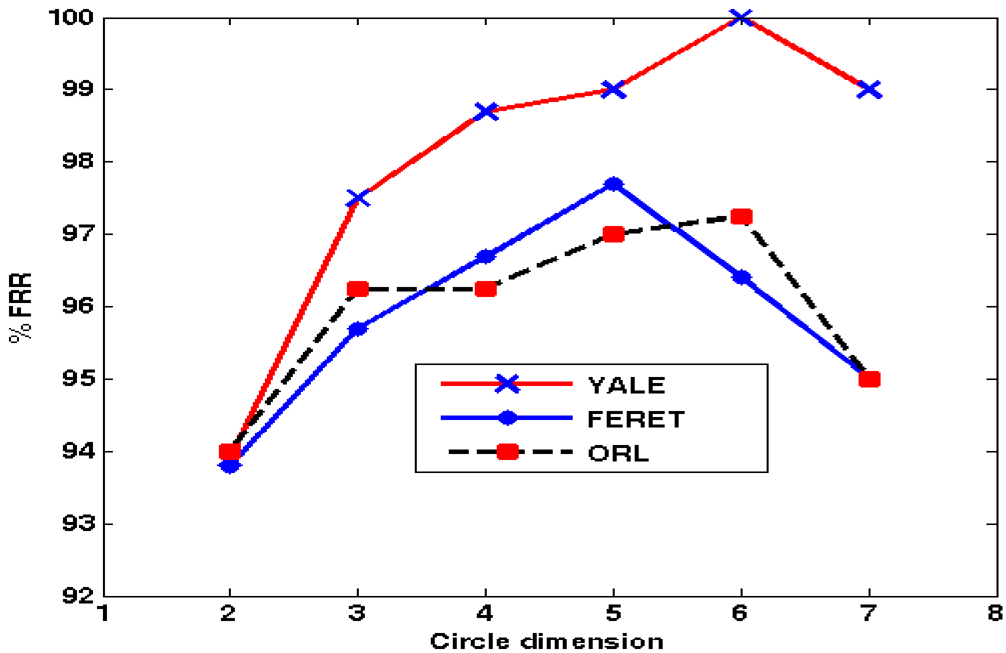


Figure 5.11. Recognition Results with constant image resolution and varying Circle Dimension

5.8.4 Experiment Set 4

In this part of experiments, image resolution through decimation is varied while keeping the feature vector dimension (radius) constant. Figure 5.12 shows the results which reflect that as the image resolution is changed the recognition rate varies. Yale database provided best recognition rate of 100% at image resolution of 56x46. Similarly experiments were also carried out on FERET and ORL database, best FRR of 97.8% and 97.5 have been noted at resolution of 128x193 and 56x46 respectively as shown in Figure 5.12.

Results tabulated in Table 5.7 illustrate that computation speed of model is improved against best FRR.

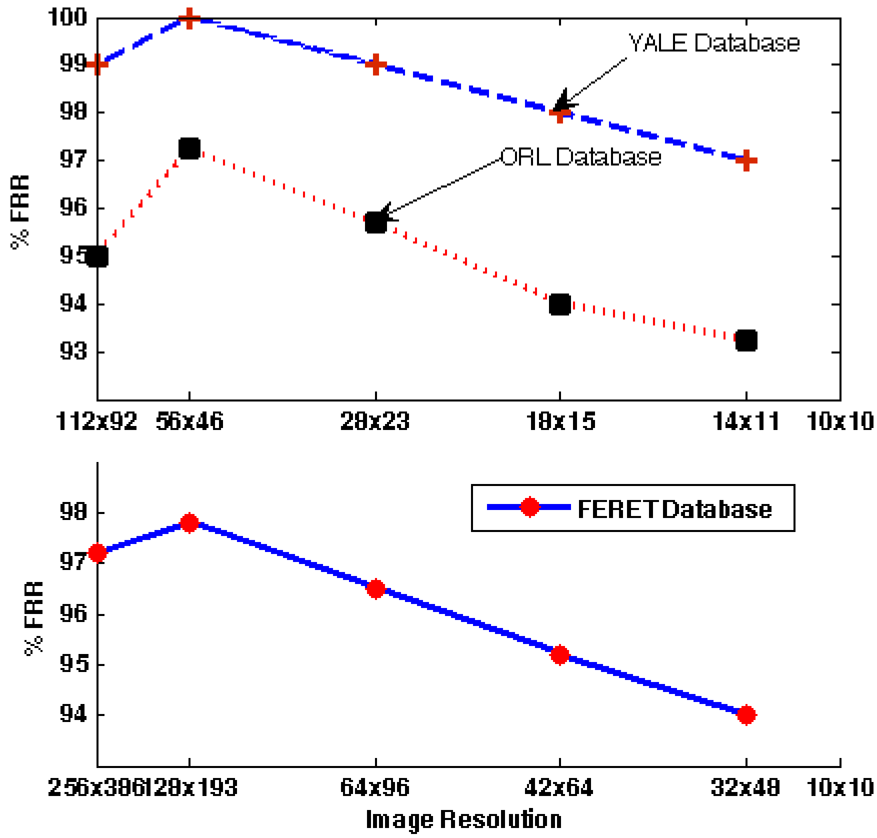


Figure 5.12. Recognition Results with varying Image resolution

Table 5.7. FRR of ORL and FERET Database with constant circle Dimension and varying image resolution

Image Resolution Variation		**Training Time Seconds**		**Recognition Time Seconds**		**FERET Database**		**ORL Database**	
ORL Database	**FERET Database**	**FERET Database**	**ORL Database**	**FERET Database**	**ORL Database**	**No of Failure**	**Error Rate(%)**	**No of Failure**	**Error Rate(%)**
112x92	**256x384**	**187.02**	**82.25**	**552.0**	**118.24**	**28**	**97.2**	**20**	**95**
56x46	**128x192**	**76.54**	**40.5**	**248.7**	**67.9**	**22**	**97.8**	**11**	**97.25**
28x23	**64x96**	**38.04**	**30.7**	**185.67**	**57.09**	**35**	**96.5**	**18**	**95.5**
18x15	**42x64**	**29.9**	**27.03**	**164.00**	**50.4**	**48**	**95.2**	**24**	**94**
14x11	**32x48**	**26.04**	**25.77**	**152.07**	**47.99**	**60**	**94**	**28**	**93**

5.9 Discrete Cosine Transform (DCT)

A discrete cosine transform (DCT) is a sinusoidal unitary transform. The DCT has been used in digital signal and image processing and particularly in transform coding systems for data compression/decompression. This type of frequency transform is real, orthogonal and separable. Algorithms for its computation have proved to be computationally efficient. In fact, the DCT has been employed as the main processing tool for data compression/decompression, image recognition and video coding standards.

The DCT is a mathematical operation that transforms a set of data which is sampled at a given sampling rate to its frequency components. The number of samples should be finite and power of two for optimal computation time. The DCT is a widely used frequency transform because it closely approximates the optimal Karhunan Loeve Transform (KLT) transform while not suffering from the drawbacks of applying the KLT. However, KLT is constructed from the eigenvalues and the corresponding eigenvectors of a covariance matrix of the data to be transformed. It is signal-dependent and there is no general algorithm for its fast computation. The DCT does not suffer from various drawbacks due to data-independent basis functions and several algorithms for fast implementation. The DCT provides a good trade-off between energy packing ability and computational complexity. A one-dimensional DCT converts an array of numbers which represent signal amplitudes at various points in time into another array of numbers each of which represents the amplitude of a certain frequency component from the original array. The resulting array contains the same number of values as the original array. The first element in the resultant array is a simple average of all the samples in the input array and is referred to as DC coefficient as shown in Figure 5.13. From the remaining elements in the resultant array each indicate the amplitude of a special frequency component of the input array and are known as AC coefficients. The frequency contents of the sample set at each frequency is calculated by taking a weighted average of the entire set. This weight coefficient is like a cosine wave whose frequency is proportional to the resultant array index.

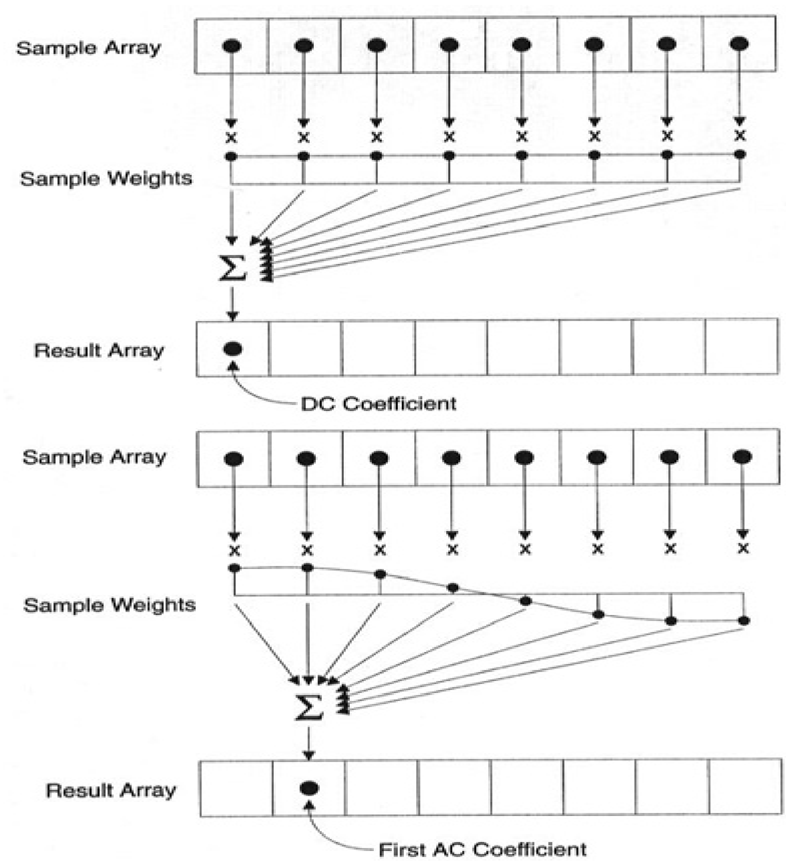


Figure 5.13. One Dimension DCT computation Process

The one-dimensional DCT can be extended to two-dimensional image arrays. The two-dimensional cosine basis functions from which sample waveforms are composed are created by multiplying a horizontally oriented set of one-dimensional 8-point basis functions by a vertically oriented set of the same functions. It logically follows that the horizontally oriented set of basis functions represent horizontal frequencies and the other set of basis functions represent vertical frequencies.

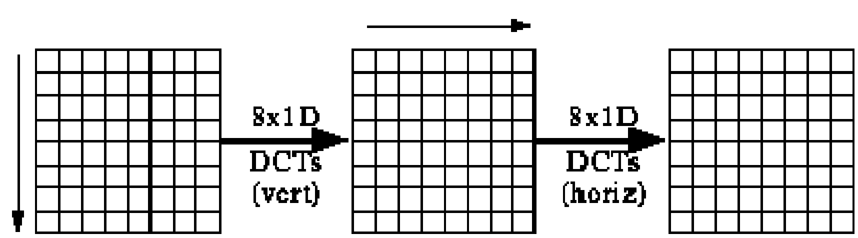


Figure.5.14. 2D –DCT computation Process

By convention the DC term of the horizontal basis functions is to the left and the DC term for the vertical basis functions is at the top. Consequently the top left element of a two-dimensional DCT matrix contains a value that is almost always of a very great magnitude. Furthermore mirroring the trend found in a one-dimensional DCT matrix the farther away an AC term is from the DC term the higher the frequency its corresponding waveform will have and the smaller its magnitude will be. The actual expression used to calculate two-dimensional DCT is shown below. The DCT helps to separate the image into parts (or spectral sub-bands) of differing importance (with respect to the image's visual quality). The DCT is similar to the discrete Fourier transform, it transforms a signal or image from the spatial domain to the frequency domain but DCT is real valued and provide a better approximation of image with fewer coefficients as compared to DFT.

The DCT for an image *f(x,y)* is:-

$$C(u,v)=\frac{2}{\sqrt{MN}}\alpha(u)\alpha(v)\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\cos\left[\frac{(2x+1)u\pi}{2M}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right] \tag{5.15}$$

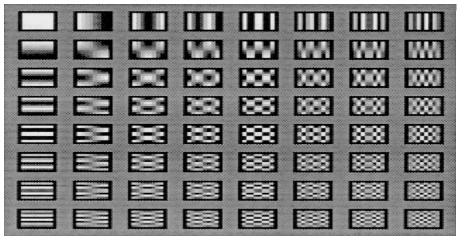
Inverse transform

$$f(x,y)=\frac{2}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\alpha(u)\alpha(v)C(u,v)\cos\left[\frac{(2x+1)u\pi}{2M}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right] \tag{5.16}$$

$$\alpha(w)=\begin{Bmatrix}\frac{1}{\sqrt{2}}\ ,if\ \ w=0\\ 1\ ,\ otherwise\end{Bmatrix}$$

5.10 DCT in Face Recognition

In proposed model of face recognition, main idea is to compute the DCT coefficients of a face image and select only a limited number of the coefficients (this limited selection corresponds to low spatial frequency coefficients) and use them as input to the recognition model. The locations of the transformed coefficients retained for each image remain unchanged from one image to another. It is computationally easier to implement and more efficient to regard the DCT as a set of basis functions which gives a known input array size (8x8) which can be pre-computed and stored. This involves computing values for a convolution mask (8 x8 window) overlap with image across all rows/columns of image. The 64 (8 x 8) DCT basis functions are illustrated in Figure 5.15.



$$\begin{bmatrix} .3536 & .3536 & .3536 & .3536 & .3536 & .3536 & .3536 & .3536 \\ .4904 & .4157 & .2778 & .0975 & -.0975 & -.2778 & -.4157 & -.4904 \\ .4619 & .1913 & -.1913 & -.4619 & -.4619 & -.1913 & .1913 & .4619 \\ .4157 & -.0975 & -.4904 & -.2778 & .2778 & .4904 & .0975 & -.4157 \\ .3536 & -.3536 & -.3536 & .3536 & .3536 & -.3536 & -.3536 & .3536 \\ .2778 & -.4904 & .0975 & .4157 & -.4157 & -.0975 & .4904 & -.2778 \\ .1913 & -.4619 & .4619 & -.1913 & -.1913 & .4619 & -.4619 & .1913 \\ .0975 & -.2778 & .4157 & -.4904 & .4904 & -.4157 & .2778 & -.0975 \end{bmatrix}$$

Figure.5.15. DCT basis functions

The DCT helps to separate the image into parts (or spectral sub-bands) of differing importance (with respect to the image's visual quality). DCT transforms the input into a linear combination of weighted basis functions. These basis functions are the frequency components of the input data. For most images, much of the signal energy lies at low frequencies (corresponding to large DCT coefficient magnitudes). These are relocated to the upper-left corner of the DCT. Conversely, the lower-right values of the DCT array represent higher frequencies and turn out to be smaller in magnitude. Horizontal frequencies increase from left to right, and vertical frequencies increase from top to bottom. The constant-valued basis function at the upper left is often called the DC basis function, and the corresponding DCT coefficient is often called the DC coefficient.

KLT-based transforms are optimal in the sense of energy compact and data decorelation. However, the KLT is data dependent and computationally expensive transform. The basis functions of DCT are data independent and its information packing ability closely approximates the optimal KLT. The generations of DCT-based feature vectors provide a good compromise between information packing ability and computational complexity. In addition to this, the DCT packs the most information into the fewest coefficients for most natural images and minimize blocking artifact. Due to this capability of DCT, it has been applied on face images and a pattern of low frequency coefficients with maximum energy is extracted for recognition purposes. This feature vector lies in upper left corner of spectrum matrix. In this research work DCT has been used to further extract the low frequency components with maximum energy from images with reduced resolution. Different dimensions of feature vector have been extracted and used for recognition.

5.11 Implementation of DCT Based Face Recognition Model

In implementation of DCT based face recognition model, following algorithm has been developed and implemented.

Begin

1. Load training images
2. Carry out image preprocessing
   - Call scale normalization module
   - Call geometric normalization module
3. Call Resolution module
4. Compute DCT Coefficients
   - Load 8x8 DCT coefficient mask
   - Compute coefficients by overlapping mask on image
   - Find out coefficients required for recognition
5. Carry out dimension reduction through selection of high energy coefficient
6. Carry out image Recognition
   - Load test image
   - Repeat the steps 2 to 5

- Carry out comparison of test image with training images through Euclidean distance
- Find out closest match

7. Display image with closest match

End

Five images of each class are used in learning process of the module. Image preprocessing as discussed in chapter 4 is carried out and image resolution variation is obtained as discussed in DFT based face recognition model. Images at specific resolution are loaded and with the help of 8 x 8 DCT basis coefficients, the DCT of complete image is computed. The result is shown in Figure 5.16.

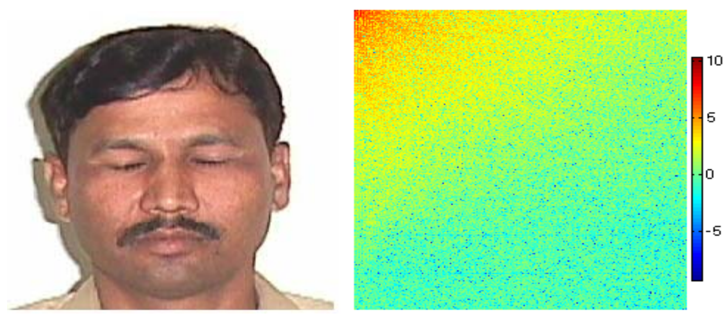


Figure 5.16. Original image and its DCT coefficient plot

Figure 5.16 reflects that high energy low frequency image components reside in upper left corner. As we move away from this corner energy is reduced. These coefficients against each image used in learning process of model are retained as feature vector. This feature vector is converted into single column vector with dimension of F x 1, where F is number of coefficients used for the recognition. In learning process of model a cosine subspace matrix of F x T is obtained where T is total number of images used for training of model. In recognition process a preprocessed test image with same image resolution as used in learning process is presented to model. DCT of test image is computed and required number of coefficients in the form of column vector are extracted. Euclidean distance is used to find out the best match. To evaluate the effect of number of DCT coefficients used for recognition on FRR, experiments have also been performed.

5.12 Experiments and Results of DCT Based Recognition Model

In proposed DCT based face recognition model two sets of experiments have been performed. In first set of experiments FRR has been obtained against varying image resolution while keeping the feature vector dimension of 64 constant. Five images of each class of all four databases discussed in chapter 4 are used for training and complete databases are used as test images.

Table 5.8. FRR of ORL and YALE Database with constant 64 feature vector Dimension and varying image resolution

Resolution Variation Factor	Image Resolution	Training Time (Seconds)		Recognition Time per Image (Seconds)		ORL Database		Yale Database	
Value of G		ORL Database	Yale Database	ORL Database	Yale Database	No of Failure	FRR (%)	No of Failure	FRR (%)
1	112x92	64.11	25.21	0.16	0.39	16	96	8	95.15
2	56x46	31.5	11.9	.078	0.189	14	96.5	7	95.75
4	28x23	30.2	7.76	.070	0.143	14	96.5	9	94.5
6	18x15	19.9	6.62	.047	0.13	14	96.5	10	93.9
8	14x11	18	5.45	.040	0.12	12	97	11	93.3

Figure 5.17 and Table 5.8 depict the contribution of resolution variation on recognition in ORL and YALE databases. Results reflect that image resolution variation has a positive contribution on FRR as ORL database provides best results of 97% at image resolution of 14x 11 and similarly YALE provides FRR of 95.75% at image resolution of 56x46. Besides this improvement in FRR, the training and recognition time of model against best FRR has also improved.

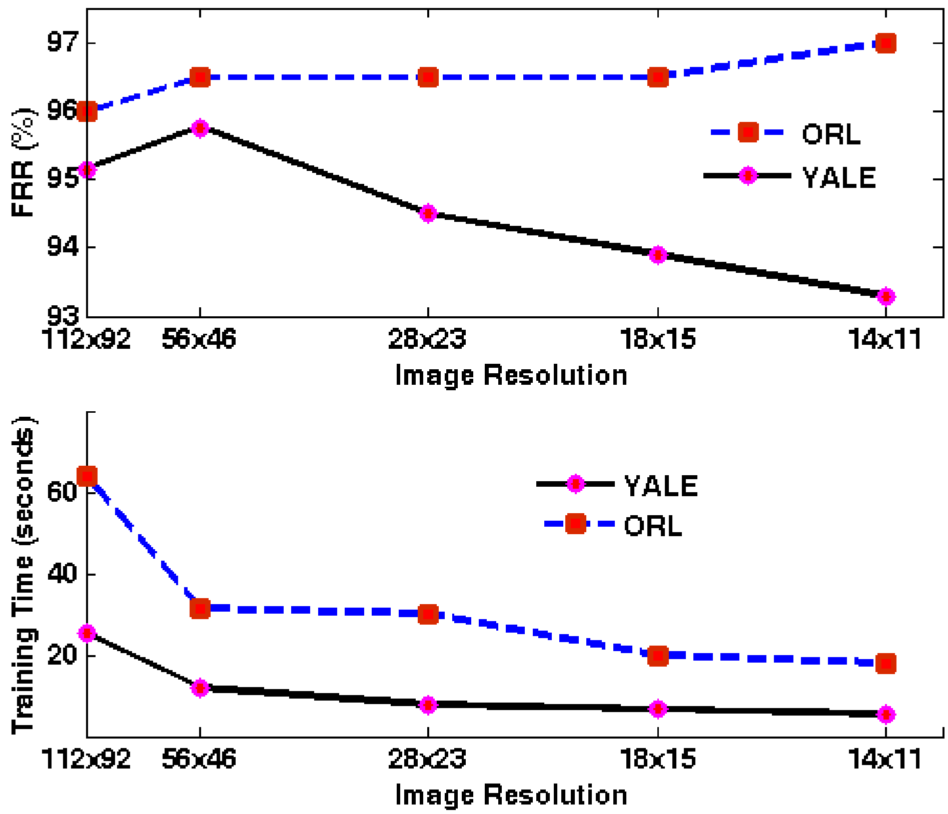


Figure.5.17. FRR of ORL and YALE database against varying image resolution

The results obtained on FERET and CEME_NUST color databases are illustrated in Table 5.9 and Figure 5.18. It is evident from results that the image resolution variation in DCT based face recognition process give improvement in FRR, but at the same time results reflect that training time of the model against best recognition results is reduced from 21.67 to 8.46 seconds against best FRR of 93.3% (CEME_NUST color database) and from 898.06 second to 106.06 against FRR of 97.3% (FERET database). Similarly, recognition time against optimum FRR is also reduced. This improvement in computation is due to dimension reduction caused by reduction in image resolution. Reduction in image resolution through image decimation process not only improves recognition results, training and recognition timings but also addresses the issue of dimension reduction in face recognition. Although application of DCT on face image and selection of 8x8 top left corner coefficients with highest energy minimizes the feature vector dimension but at the same time resolution reduction decreases the time consumption required for computation of DCT spectrum.

Table 5.9. FRR of FERET and CEME_NUST color Database with constant 8x8 feature vector Dimension and varying image resolution

Image Resolution		**Training Time (Seconds)**		**Recognition Time per Image (Seconds)**		**FERET Database**		**CEME_NUST Color Database**	
FERET Database	**CEME_NUST Color Database**	**FERET Database**	**CEME_NUST Color Database**	**FERET Database**	**CEME_NUST Color Database**	**No of Failure**	**FRR (%)**	**No of Failure**	**FRR (%)**
256x384	**112x100**	**898.06**	**21.67**	**0.89**	**0.199**	**86**	**89.16**	**13**	**89.16**
128x192	**56x50**	**722.6**	**10.37**	**0.59**	**0.107**	**83**	**88.33**	**14**	**88.33**
64x96	**28x25**	**106.06**	**8.46**	**.226**	**.071**	**27**	**97.3**	**8**	**93.33**
42x64	**18x16**	**75.56**	**6.75**	**.194**	**.063**	**28**	**97.2**	**11**	**90.80**
32x48	**15x12**	**60.07**	**5.16.**	**.176**	**.058**	**31**	**96.9**	**11**	**90.80**

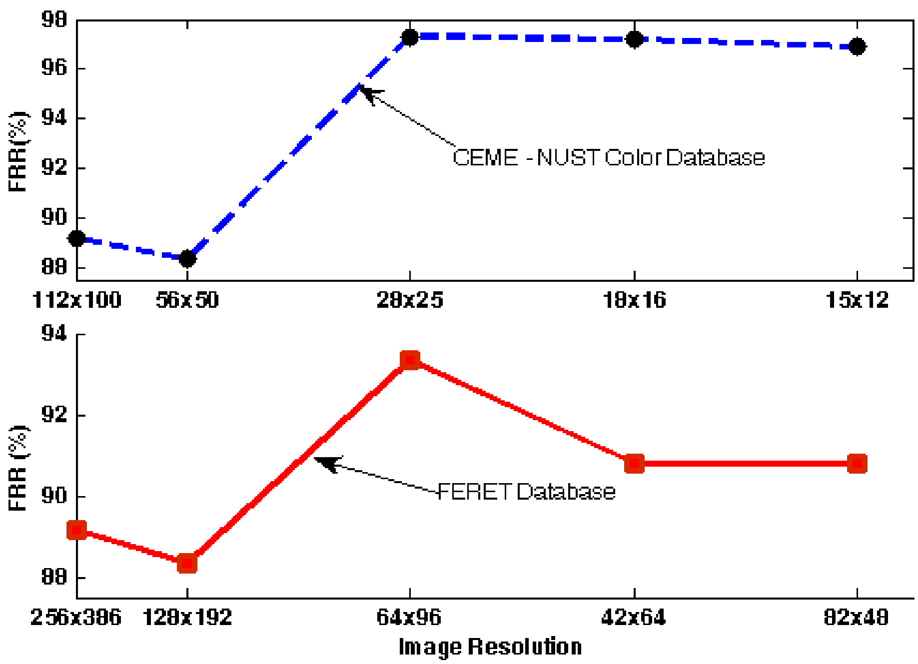


Figure.5.18. FRR of CEME_NUST and FERET database against varying image resolution

In second set of experiments results have been obtained by varying the number of DCT coefficients retained for recognition.

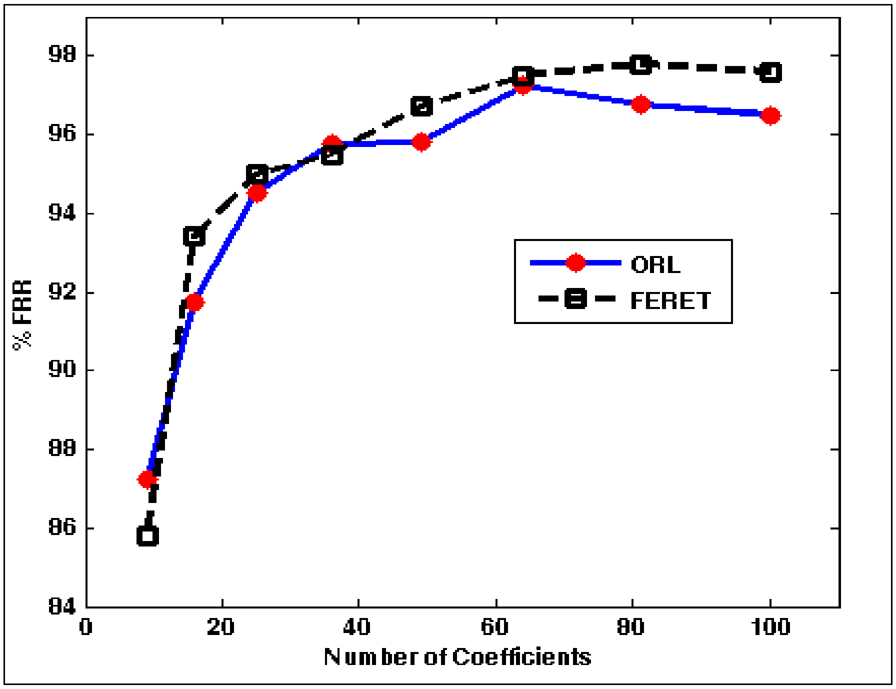


Figure.5.19. FRR of ORL and FERET database against varying feature vector dimensions

Figure 5.19 depicts that the results gathered on ORL and FERET databases with varying number of DCT coefficients used for recognition. Once the feature vector dimension is increased to 8x8 (64 coefficients), the recognition rate is decreasing in ORL database but in FERET database this fall starts once feature vector dimension is increased beyond 9x9 (81 coefficients). This conclusion illustrates that low frequency components lie in upper left corner of DCT spectrum and for each database a particular dimension of feature vector is required. This feature vector represents the facial information required for recognition in a compact fashion without any redundancy in data used for optimum recognition rate as shown in Figure 5.19.

5.13 Facial Multiresolution Analysis through Wavelets

Wavelet transform techniques are not too old and these techniques are being used in modern signal and image processing including multiresolution analysis, sound synthesis, computer vision, graphics and image compression [108]. The transform of a signal is just another form of representing the signal. It does not change the information content present in the signal. The Wavelet Transform provides a time-frequency representation of the signal. It was developed to overcome the short coming of the Short Time Fourier Transform (STFT), which can also be used to analyze non-stationary signals. While STFT gives a constant resolution at all frequencies, the Wavelet Transform uses multi-resolution technique by which different frequencies are analyzed with different resolutions. A wave is an oscillating function of time or space and is periodic as shown in Figure 5.20(a). In contrast, wavelets shown in Figure 5.20(b) are localized waves. They have their energy concentrated in time or space and are suited for analysis of transient signals. Fourier Transform and STFT use waves to analyze signals, whereas the Wavelet Transform uses wavelets of finite energy

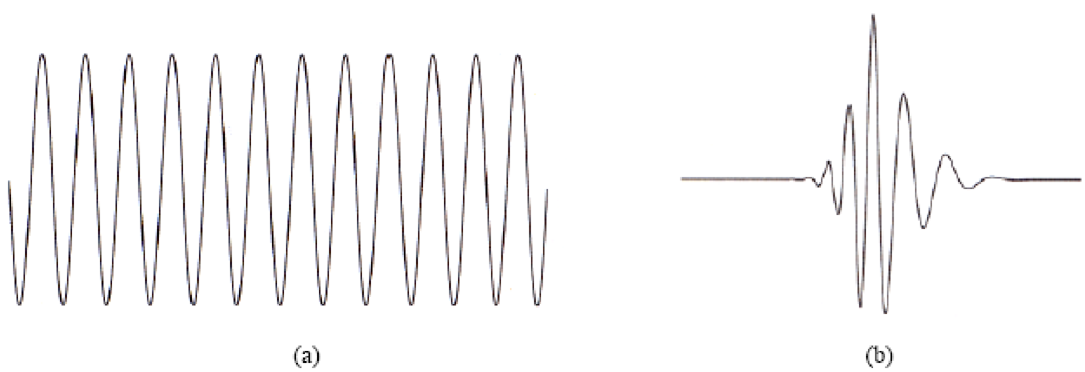


Figure.5.20 (a) a wave and (b) Wavelet

The Continuous Wavelet Transform (CWT) is provided by Equation 5.14, where x(t) is the signal to be analyzed and $\psi(t)$ is the mother wavelet or the basis function. All the wavelet functions used in the transformation are derived from the mother wavelet through translation (shifting) and scaling (dilation or compression).

$$X_{WT}(\tau,s)=\frac{1}{\sqrt{|s|}}\int x(t).\psi^{*}\left(\frac{t-\tau}{s}\right)dt \tag{5.17}$$

The Wavelet Series is obtained by discretizing CWT. This aids in computation of CWT using computers and is obtained by sampling the time-scale plane. The sampling rate can be changed accordingly with scale change without violating the Nyquist criterion. Nyquist criterion states that, the minimum sampling rate that allows reconstruction of the original signal is 2ω radians, where ω is the highest frequency in the signal. Therefore, as the scale goes higher (lower frequencies), the sampling rate can be decreased thus reducing the number of computations. The Wavelet Series is just a sampled version of CWT and its computation may consume significant amount of time and resources depending on the resolution required.

5.14 Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT), which is based on sub-band coding, is found to yield a fast computation of Wavelet Transform. It is easy to implement and reduces the computation time and resources required. The foundations of DWT go back to 1976 when techniques to decompose discrete time signals were devised. Similar work was done in speech signal coding which was named as sub-band coding. In 1983, a technique similar to sub-band coding was developed which was named pyramidal coding. Later many improvements were made to these coding schemes which resulted in efficient multi-resolution analysis schemes. In CWT, the signals are analyzed using a set of basis functions which relate to each other by simple scaling and translation. In the case of DWT, a time-scale representation of the digital signal is obtained using digital filtering techniques. The signal to be analyzed is passed through filters with different cutoff frequencies at different scales.

Filters are one of the most widely used signal processing functions. Wavelets can be realized by iteration of filters with rescaling. The resolution of the signal which is a measure of the amount of detail information in the signal is determined by the filtering operations and the scale is determined by upsampling and downsampling (subsampling) operations. The DWT is computed by successive lowpass and highpass filtering of the discrete time-domain signal as shown in Figure 5.21.

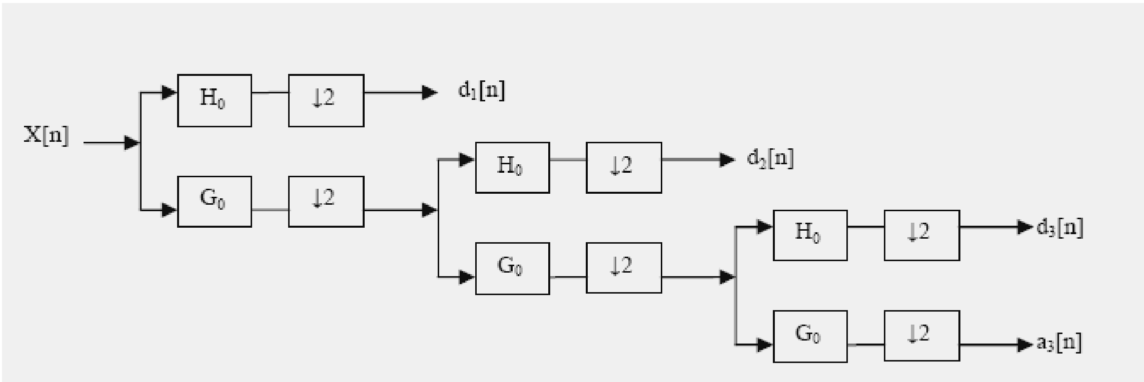


Figure. 5.21. DWT decomposition process

At each decomposition level, the half band filters produce signals spanning only half the frequency band. This doubles the frequency resolution as the uncertainty in frequency is reduced by half. In accordance with Nyquist's rule if the original signal has a highest frequency of ω which requires a sampling frequency of 2ω radians then it now has a highest frequency of ω/2 radians. It can now be sampled at a frequency of ω radians thus discarding half the samples with no loss of information. This decimation by 2 halves the time resolution as the entire signal is now represented by only half the number of samples. Thus while the half band low pass filtering removes half of the frequencies and thus halves the resolution, the decimation by 2 doubles the scale.

In two dimensional wavelet transform a two dimensional scaling factor $\varphi(x, y)$ and three two dimensional wavelets, $\psi^H(x, y), \psi^V(x, y) \, and \, \psi^D(x, y)$ are required. Each is the product of a one dimensional scaling function φ and corresponding wavelet ψ. The separable scaling function is

$$\varphi(x, y) = \varphi(x)\varphi(y) \tag{5.18}$$

and the separable directionally sensitive wavelets are

$$\psi^H(x, y) = \psi(x)\psi(y) \tag{5.19}$$

$$\psi^V(x, y) = \psi(x)\psi(y) \tag{5.20}$$

$$\psi^D(x,y) = \psi(x)\psi(y) \tag{5.21}$$

These wavelets measure functional variations i.e, intensity or gray level variations for images along different directions:- ψ^H measures variations along columns (for example, horizontal edges), ψ^V responds to variations along rows (like vertical edges) and ψ^D corresponds to variations along diagonals.
The discrete wavelet transform of a function *f(x,y)* of size *M* x *N* is:-

$$W_\varphi\left(j_o,m,n\right) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)\varphi_{j_o,m,n}(x,y) \tag{5.22}$$

$$\overset{i}{W_\psi}\left(j,m,n\right) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)\psi^i{}_{j,m,n}(x,y) \tag{5.23}$$

$W_\varphi\left(j_o,m,n\right)$ coefficients define an approximation of *f(x,y)* at scale j_o. $\overset{i}{W_\psi}\left(j,m,n\right)$ coefficients add horizontal , vertical and diagonal details for scale $j \le j_o$. Similarly the *f(x,y)* can be obtained via inverse DWT.

$$\begin{aligned} f(x,y) = {} & \frac{1}{MN} \sum_m \sum_n W_\varphi\left(j_o,m,n\right)\varphi_{j_o,m,n}(x,y) \\ & + \frac{1}{MN} \sum_{i=H,V,D} \sum_{j=j_o}^{\infty} \sum_m \sum_n W_\psi^i(j,m,n)\psi^i_{j,m,n}(x,y) \end{aligned} \tag{5.24}$$

Wavelet transform techniques achieve optimal decomposition without affecting much the image quality. At the same time wavelet transform and wavelet packet analysis have provided a new subspace for image recognition. Fotlyniewiez [109] proposed an automatic face recognition using nonlinear filtering to enhance intrinsic features of face and used a high order neural network classifier for training and recognition of faces. Lee and Chung [110] employed the wavelet-based Fisher Linear Discriminant (FLD) recognition process. Zhu and Orchad [111] captured local discriminative features in the space frequency domain for face detection using wavelet packet analysis. Ma and Tang [112] used discrete wavelet face graph matching approach for the purpose. Liu [113] used Haar wavelet for effective human face detection. Yang et al [114] presented an application of nonlinear wavelet approximation to recognize faces and the advantages of nonlinear wavelet approximation compared to its linear counterpart. L.Wiskott [115] used labeled graph based on Gabor wavelet transform for face recognition application.

In face recognition the wavelets provide best multiresolution analysis of face image and segregate the low frequency components of image required for better recognition. At the same time wavelets also carry out the image dimension reduction which enhances the computation speed and decreases improved hardware reliance. In this model of face recognition effects of image resolution on face recognition are evaluated. Image resolution is varied through decimation algorithm which provides best image resolution for better recognition.

5.15 Wavelet Application for Multiresolution Analysis and Dimension Reduction

Wavelets are functions that satisfy certain mathematical requirements and are used in representing data or other functions. Like sines and cosines in Fourier transform, wavelets are used as basis functions in representing other functions. However unlike the Fourier transform the wavelet transform processes data at different scales or resolutions. Wavelets are formed by dilation and translations of a signal function $\psi(x)$ called mother wavelet. Wavelet signal processing involves wavelet analysis (decomposition) and wavelet synthesis (reconstruction). The work of Daubechies reflects that digital filter banks can be used under certain conditions to implement fast wavelet transforms. The digital filter bank method filters a signal by several parallel filters followed by sub-sampling. A two channel filter bank has a low pass filter L and a high pass filter H. L removes high frequencies and produces low frequency components of signal (approximation). H removes the low frequencies and generates high frequency components of the signals (details). These filters separate the signal into frequency bands, where the signal length becomes double. To correct this problem down sampling is used. Only half of output components of each filter are kept. Now this decomposition process can be iterated. If successive approximations are decomposed, it generates the wavelet decomposition tree. Wavelet synthesis process involves up sampling and filtering. In up sampling the length of signal is doubled by adding zeros between components. The low pass and high pass decomposition filters together with their reconstruction filters form a system called quadrature mirror filters. Aliasing and distortion can be removed by selecting proper decomposition and reconstruction orthogonal or biorthogonal filters.

The two dimensional wavelet transform can be derived from the one dimensional wavelet transform. After one step four sub-bands emerge (one is averaging image $I_{LL,}$ and three detail images I_{LH}, I_{HL} and $I_{HH)}$. The DWT has been used for texture classification [116] and image compression [117] due to multiresolution decomposition property. The wavelet decomposition technique was also used to extract the intrinsic features for face recognition. In wavelets packet analysis both the high and low frequency filters are iterated but in wavelet transform only the low pass filter is iterated where it is assumed that low frequency contents contribute more than the higher frequencies to represent information in face images.

The DWT of a signal *x* is calculated by passing it through a series of filters. First the samples are passed through a low pass filter with impulse response *g* resulting in a convolution of the two:

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k].g[n-k] \tag{5.25}$$

The signal is also decomposed simultaneously using a high-pass filter *h*. The outputs give the detail coefficients (from the high-pass filter) and approximation coefficients (from the low-pass). It is important that the two filters are related to each other and they are known as a quadrature mirror filter.

However, since half the frequencies of the signal have now been removed, half the samples can be discarded according to Nyquist's rule. The filter outputs are then downsampled by 2:

$$\begin{bmatrix} y_{low}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n-k] \\ y_{high}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n-k] \end{bmatrix} \tag{5.26}$$

This decomposition has halved the time resolution since only half of each filter output characterises the signal. However, each output has half the frequency band of the input so the frequency resolution has been doubled.

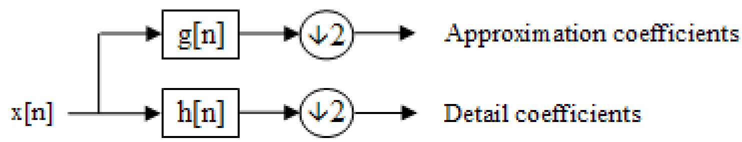


Figure 2.22. Block diagram of filter analysis

With the downsampling operator $\downarrow$

$$(y \downarrow k)[n] = y[k.n] \tag{5.27}$$

The above summation can be written more concisely.

$$\begin{bmatrix} y_{low} = (x * g) \downarrow 2 \\ y_{high} = (h * g) \downarrow 2 \end{bmatrix} \tag{5.28}$$

However computing a complete convolution $x * g$ with subsequent downsampling would waste computation time. The Lifting scheme is an optimization where these two computations are interleaved.

Same procedure can be extended to two-dimensional (2-D) images. DWT on image x[m,n] can be performed by consecutively applying one-dimensional wavelet transform to the rows and columns of the two-dimensional data.

$$DWT_n[DWT_m[x[m,n]]] \tag{5.29}$$

Two-dimensional WT decomposes an image into “subbands” that are localized in frequency and orientation. A image is passed through a series of filter bank stages. The high-pass filter (wavelet function) and low-pass filter (scaling function) are finite impulse response filters. In other words, the output at each point depends only on a finite portion of the input. The filtered outputs are then down sampled by a factor of 2 in the horizontal direction. These signals are then filtered by an identical filter pair in the vertical direction. We end up with a decomposition of the image into 4 subbands denoted by LL, HL, LH, HH. Each of these subbands can be thought of as a smaller version of the image representing different image properties. The band LL is a coarser approximation to the original image. The bands LH and HL record the changes of the image along horizontal and vertical directions, respectively. The HH band shows the high frequency component of the image. Second level decomposition can then be conducted on the LL subband. Under frequency-based representation, only high-frequency spectrum is affected, called high-frequency phenomenon. Moreover, changes in pose or scale of a face affect the intensity manifold globally, in which only their low-frequency spectrum is affected, called low-frequency phenomenon. Only a change in face will affect all frequency components.

5.16 Classification of Wavelets

Wavelets can be classified into two classes: (a) orthogonal and (b) biorthogonal. Based on the application, either of them can be used.

5.16.1 Features of Orthogonal Wavelet Filter Banks

The coefficients of orthogonal filters are real numbers. The filters are of the same length and are not symmetric. The low pass filter G_0 and the high pass filter H_0 are related to each other by

$$H_0 (z) = z^{-N} G_0 (-z)^{-1} \tag{5.30}$$

The two filters are alternated flip of each other. The alternating flip automatically gives double-shift orthogonality between the lowpass and highpass filters (i.e., the scalar product of the filters, for a shift by two is zero). Filters that satisfy Equation 5.20 are known as Conjugate Mirror Filters (CMF). Perfect reconstruction is possible with alternating flip. Also, for perfect reconstruction, the synthesis filters are identical to the analysis filters except for a time reversal. Orthogonal filters offer a high number of vanishing moments. This property is useful in many signal and image processing applications. They have regular structure which leads to easy implementation and scalable architecture.

5.16.2 Features of Biorthogonal Wavelet Filter Banks

In the case of the biorthogonal wavelet filters, the low pass and the high pass filters do not have the same length. The low pass filter is always symmetric, while the high pass filter could be either symmetric or anti-symmetric. The coefficients of the filters are either real numbers or integers. For perfect reconstruction, biorthogonal filter bank has all odd length or all even length filters. The two analysis filters can be symmetric with odd length or one symmetric and the other antisymmetric with even length. Also, the two sets of analysis and synthesis filters must be dual. The linear phase biorthogonal filters are the most popular filters for data compression applications.

5.17 Face Recognition Using Wavelets

In recent developments of recognition, wavelet transform has been employed for face recognition. For 2D discrete wavelet transform (DWT), an image is represented in terms of translations and dilations of a scaling function and a wavelet functions.

The scaling and wavelet coefficients can be easily computed using a 2D-filter bank consisting of low-pass and high-pass filters. After one level of 2D decomposition, an image is divided into four sub-bands: LL (Low-Low), which is generated by the approximation coefficients; LH (Low-High), HL (High-Low), and HH (High-High), which are generated by the detail coefficients, as shown in Figure 5.21.

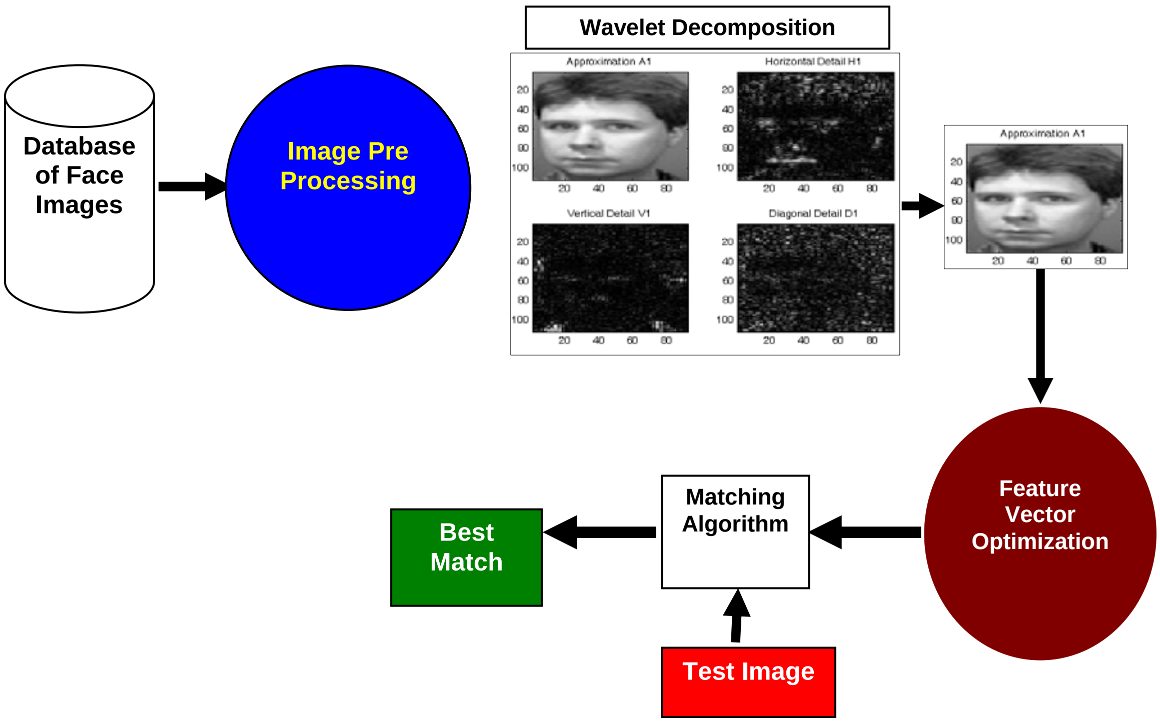
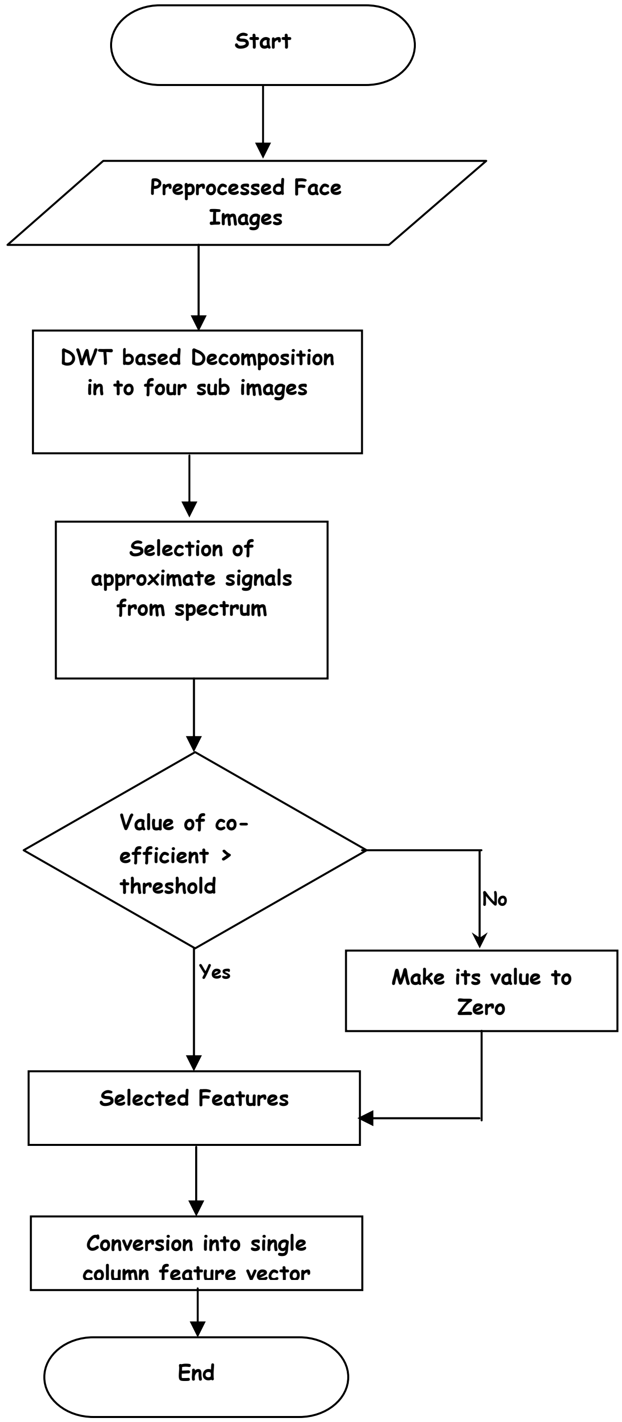


Figure.5.23. Face Recognition Model

In case of face recognition most of the information required for recognition lies in low frequency components of the image. So trend signals of image transformed through first level DWT are used for the recognition purpose as shown in Figure 5.23.

5.18 Implementation

The idea behind using discrete wavelet transform is to analyze the signal at different scales or resolutions, which is called multiresolution as wavelets are a class of functions used to localize a given signal in both space and scaling domains. In all face recognition problems, information in low spatial frequency bands play a dominant role in face recognition. Moreover, it is evident from experiments that facial expressions and small occlusions affect the intensity manifold locally. Under frequency-based representation, only high-frequency spectrum is affected, called high-frequency phenomenon. Likewise, changes in pose or scale of a face affect the intensity manifold globally, in which only their low-frequency spectrum is affected called low-frequency phenomenon. So to use this global part of facial spectrum the face images are applied by discrete wavelet transform.



Only the approximate part of this transformed spectrum is taken as feature vector to represent the facial information of the image required for recognition as shown in flow chart. Scale and geometric normalization of all images of database is carried out and their background is made uniform through morphological operations. These pre-processed face images are taken as input to the model and their DWT spectrum is computed. Approximate portion out of four images of different frequency bands is selected. Coefficient optimization is carried out by making the value of all coefficients below a certain threshold to zero. This feature vector is converted into a single column vector. Single column vector of all the images used in training of model is obtained. Finally, a matrix of L x T dimension is obtained where L is total number of coefficients retained to represent the facial information of an image and T is total number of images used in the training of model. In recognition process, a preprocessed test image is presented to the model. Its DWT spectrum is computed and feature selection is done. A feature vector of L x 1 dimension is retained as facial representation of image. This vector is compared with each column of matrix obtained during training of the model. For classification of the image, a dissimilarity matrix is obtained which is converted into single row matrix where each column reflects the accumulative difference of training image with the test image. The minimum accumulative difference is taken as best match against test image.

5.19 Choice of Wavelet Family for Recognition

Five images from each class are used for the training of the model and complete database is presented to proposed model as test image. First to find out the best wavelet for recognition purposes results, have been gathered by applying Daubechie, Biorthogonal, Symlet, and coiflets DWTs on preprocessed images of ORL and YALE databases without carrying out coefficient optimization of wavelet coefficients and variation in image resolution. Table 5.10 shows the recognition results of Daubechies (db1 to db6) and Bioorthgal (Bior1.3 to Bior5.5) wavelet families. Table 5.11 reflects the recognition performance of Symlet (sym1 to sym7) and Coiflets (coif1 to coif7) DWT on ORL and YALE databases. Result shown in Table 5.10 and 5.11 revealed that DWT symlets4 with decomposition level one provides best recognition results on preprocessed images.

Table 5.10. Recognition Results of Daubechies and Biorthogonal Family DWT on ORL and Yale datasets

Wavelet Family	Recognition Rate %		Wavelet Family	Recognition Rate %	
Daubechies	ORL Dataset	Yale Dataset	Biorthogonal	ORL Dataset	Yale Dataset
db1	96	93	Bior1.3	96.5	94
db2	96.5	93.5	Bior 2.2	96	93
db3	96	94	Bior 2.6	95	93.5
db4	97	92	Bior 3.9	94	93.5
db5	94	92.5	Bior 4.4	95.5	92
db6	94	94	Bior 5.5	93	93

Table 5.11. Recognition Results of Symlet and Coiflets Family DWT on ORL and Yale datasets

Wavelet Family	Recognition Rate %		Wavelet Family	Recognition Rate %	
Symlets	ORL Dataset	Yale Dataset	Coiflets	ORL Dataset	Yale Dataset
sym1	96	94	Coif1	97	94
sym2	96.5	94	Coif 2	97	92
sym3	95	93	Coif 3	96	91
sym4	97.5	94.5	Coif 4	96	92
sym5	95	94	Coif 5	96	92
sym6	94	93	-	-	-
sym7	94	93	-	-	-

Symlets wavelet was proposed by Daubechies as modifications to the db family. The properties of the two wavelet families are mostly similar as shown in Table 5.12. Here the symmetrical, orthogonal and biorthogonal properties of symlets wavelet are exploited to obtain the low frequency image components which provide best face image recognition.

Table 5.12. Few Properties of Haar, Daubechies, Symlet and Biorthogonal Wavelets

Property	**Haar**	**Daubechies**	**Symlet**	**Biothogonal**
Orthogonal	Yes	Yes	Yes	No
Biorthogonal	Yes	Yes	Yes	Yes
Compact support	Yes	Yes	Yes	Yes
DWT	Possible	Possible	Possible	Possible
CWT	Possible	Possible	Possible	Possible
Filter length	2	2N-1	2N-1	$2Nr+1, 2^{Nd}+1$
Symmetry	Yes	Far from	Near from	Yes

5.20 Feature Vector Optimization

As not all the coefficients of a wavelet transform have the information required for recognition, coefficient optimization is carried out by defining threshold value. This threshold value is defined in such a way that image quality and coefficients required for recognition are not compromised. The coefficients below this threshold are made zero which help in reducing overall computational burden. Plot in Figure 5.24 demonstrates the effect of threshold on wavelet coefficients. The value of $C_{i,j}$ below threshold is replaced with zero, whereas the value above or equal to threshold is retained.

$$
\begin{array}{l}
\text{if} \\
C_{i,j} < Th \\
\quad C_{i,j} = 0 \\
else \\
C_{i,j} = C_{i,j} \\
where \\
C_{i,j} \;\; is\ value\ of\ cA\ spectrum\ at\ i, j\ location \\
Th\ is\ defined\ value\ of\ threshold
\end{array}
$$

5.21 Experiments and Results with Better Choice of Coefficient and Wavelet

In second set of experiment Symlet4 (level one) has been applied on preprocessed images with varying resolution. Experiments on ORL, CEME_NUST color, YALE and FERET datasets have been carried out and it was found that each dataset at a specific resolution provides best recognition results. In all the tests, five images of each individual have been used for training purpose and complete database has been used for recognition. Results are shown in Figure 5.25 and Table 5.13.

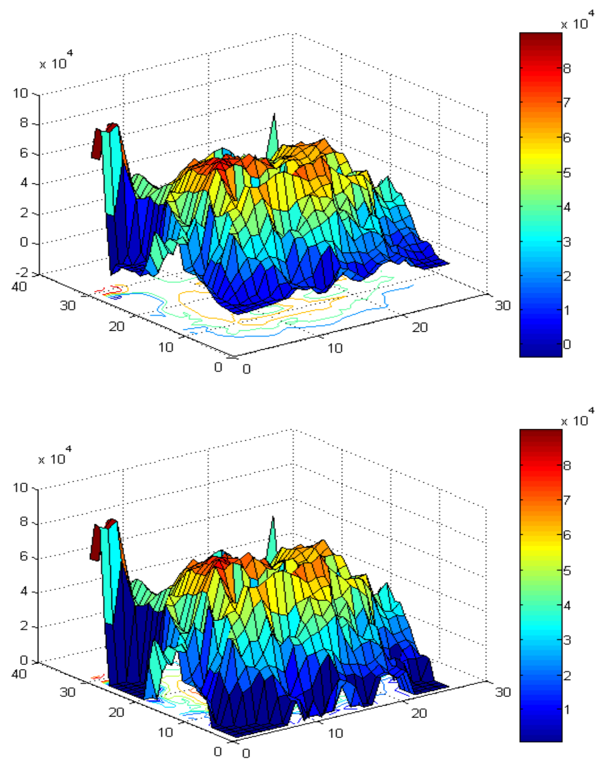


Figure 5.24. Wavelet Coefficients before (Upper) and after applying threshold (Lower)

Table 5.13. FRR of ORL and YALE Databases with Varying Image Resolution

Resolution Variation Factor	**Training Time (Seconds)**		**Recognition Time per Image (Seconds)**		**ORL Database**		**YALE Database**	
Value of G	**ORL Database**	**YALE Database**	**ORL Database**	**YALE Database**	**No of Failure**	**FRR (%)**	**No of Failure**	**FRR (%)**
112x92	60.3	24.5	0.53	0.38	8	98	5	96.9
56x46	38.7	12.76	0.43	0.189	4	99	2	98.7
28x23	27.04	9.64	0.25	0.143	7	98.25	4	97.5
18x15	21.8	9.02	0.24	0.130	15	96.25	9	94.5
14x11	20.07	8.39	0.22	0.121	16	96	9	94.5

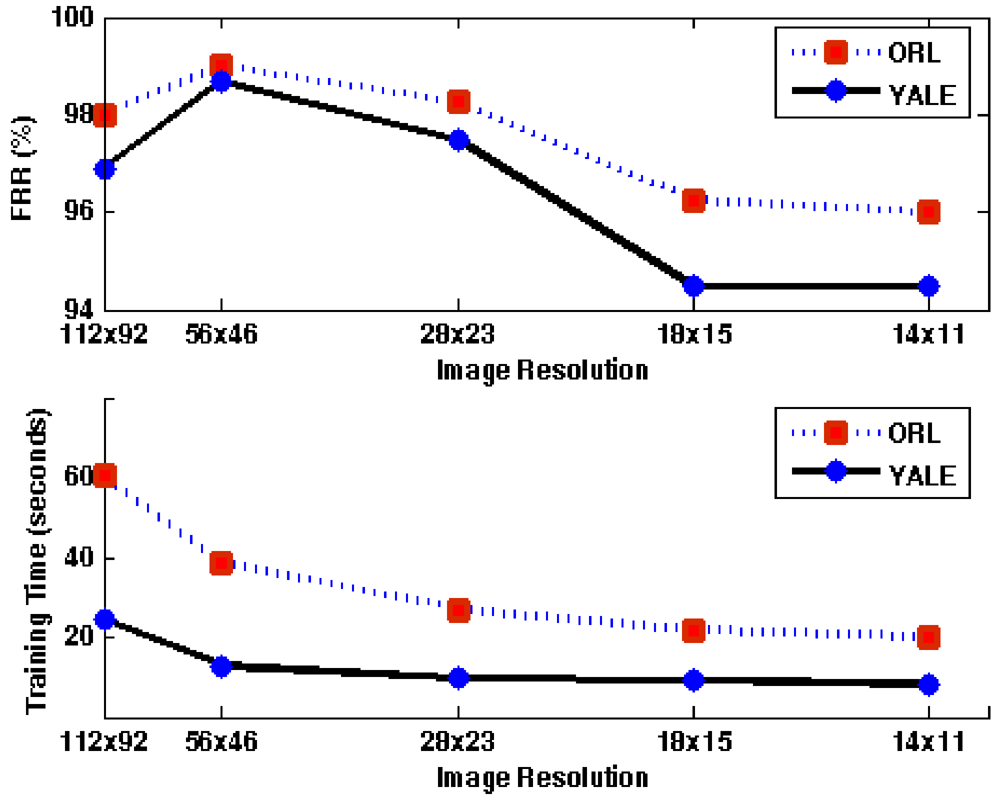


Figure. 5.25. Results of ORL and Yale database with Varying resolution level

The results gathered on ORL and Yale databases show that ORL database provides best FRR of 99% and YALE database achieves optimum FRR of 98.7% at image resolution of 56x46. In wavelet based face recognition model curse of dimensionality is addressed in two fold: first image dimension is reduced through image resolution reduction and second Symlet4 DWT is applied on face image and portion of spectrum containing low frequency is taken for further processing.

Results in Figure 5.26 and Table 5.14 reflect optimum FRR of 98.3% and 98% of CEME_NUST color database and FERET database at image resolution of 56x50 and 128x193 respectively. Table 5.14 also highlights the reduction of model training and recognition timings against best FRR.

Table 5.14. FRR of FERET and CEME_NUST color Database with Varying Image Resolution

Image Resolution		Training Time (Seconds)		Recognition Time per Image (Seconds)		FERET Database		CEME_NUST Color Database	
FERET Database	CEME_NUST Color Database	FERET Database	CEME_NUST Color Database	FERET Database	CEME_NUST Color Database	No of Failure	FRR (%)	No of Failure	FRR (%)
256x384	112x100	126.7	22.08	0.46	0.204	59	94.1	29	75.2
128x192	56x50	40.14	11.9	0.22	0.072	22	97.8	4	96.6
64x96	28x25	27.23	6.76	0.166	0.045	21	98	2	98.3
42x64	18x16	16.29	5.67	0.146	0.039	24	97.6	3	97.5
32x48	15x12	11.672	3.14	0.138	0.035	25	97.5	3	97.5

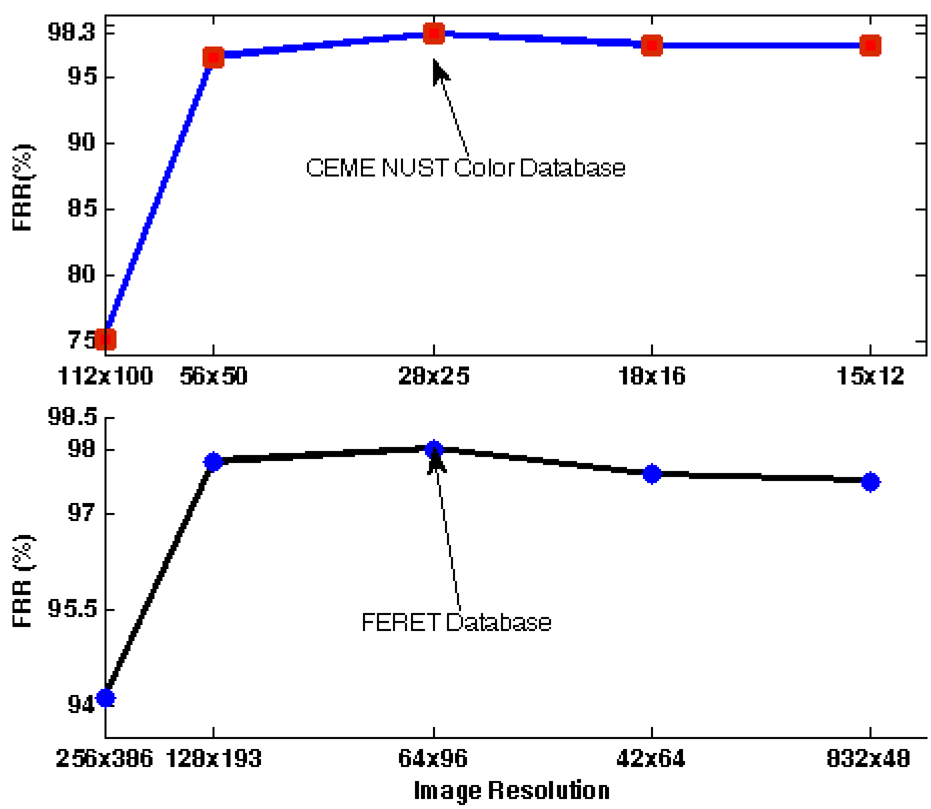


Figure. 5.26. Results of FERET and CEME_NUST Color database with Varying image Resolution.

5.22 Discussion of Frequency Based Face Recognition Model

A known problem in performing face recognition is to determine what facial features constitute the most relevant dimensions of a face. In attempting to solve this problem, there has been controversy with respect to the relevance of various spatial frequency bandwidths for face perception and processing. Structure and internal geometry of face images describes that only low-frequency spectrum is sufficient for the processing of faces for recognition. Whereas high frequency components of facial image convey facial information which does not contribute much in recognition process.

As human face is a non-rigid object, it has abundant facial expressions and these expressions influence the local spatial components of face. In this research work, the relationship between variations in facial appearance and their deformation spectrum have been investigated and it has been found that facial expressions and small occlusions affect the intensity manifold locally. The quality of the reconstructed image is very good if the image is restored with the lower half-frequency spectrum. Moreover, changes in pose or scale of a face affect the intensity manifold globally in which only their low frequency spectrum is altered called low-frequency phenomenon. Only a change in face will influence all frequency components. So, different facial expressions can be minimized by removing the high-frequency components.

In DFT based face recognition model, two distinct methods of feature extraction are used. Before applying transformation to face images, their resolution was varied through decimation process and a Gaussian pyramid of varying image resolution was obtained. The model has been tested on all images of Gaussian pyramid by keeping the image resolution constant in one set of experiments. Figures 5.10 and 5.12 reflect that recognition rate is affected by changing the image resolution. This change in recognition rate is positive till the time resolution is varied up to a certain limit beyond which change is negative. This phenomenon established the conjecture that face recognition success rate will be maximum at a specific image resolution. But at the same time it must satisfy the following condition:

$$D_{axb} \in I_{mxn}$$
$$S_{kxk} \in D_{axb}$$
$$C_{ixj} \in D_{axb} \quad \textit{and it must satisfy}$$
$$0 < Dim\ (S_{kxk}\ or\ C_{ixj}) \leq min(Dim(f))$$

where
D_{axb} *is decimated image with reduced image resolution*
S_{kxk} *is feature vector with square selection method*
C_{ixj} *is feature vector with circular selection method*
$min(Dim(f))$ *is* minimum dimension *of face image required for best facial feature representation with improved recognition*

To achieve S_{kxk} or C_{ixj} close to *Min(Dim(f))* is the main problem in any face recognition model. In proposed frequency based face recognition models it is achieved in two steps:

- By obtaining best image resolution required for recognition through decimation process
- Using two distinct better feature extraction methods

First step is carried out by changing the face image resolution through decimation. Image decimation process has been explained in preceding sections where first averaging (low pass) filter is applied on the image which extracts the low frequency components of the image by reducing the facial expression changes in the output image. Later on image decimation is applied whose order or down scale factor depends upon the order of low pass filter. As the decimation down scale factor is increased, resultant image size and resolution is decreased by reaching at a specific level where it gives best recognition rate. Further increase in decimation results in reduction of required image information and accordingly reduction in FRR. Second step is accomplished by using square and circular feature extraction methodologies.

In this research work the relation between facial expression compensation within same class and low frequency components of face image has also been investigated. Image derivative is used to highlight the facial expression in a face image and its following fundamental properties have been exploited to explain the facial expressions compensation through varying image resolution.

- First derivative must be zero in areas of constant gray levels.
- First derivative must be non zero at onset of gray level step or ramp.
- First derivative must be non zero along ramps.
- First derivative must be zero along ramps of constant slope.

A definition of first order and second order derivative of one dimensional function *f(x)* is given in expression 5.24 which can be extended to two dimensions.

$$\begin{bmatrix} \dfrac{\partial f}{\partial x} = f(x+1) - f(x) \\ \dfrac{\partial^2 f}{\partial^2 x} = f(x+1) + f(x-1) - 2f(x) \end{bmatrix} \tag{5.31}$$

A high pass filter shown in Figure 5 based on image derivative properties is used to make the facial expressions in face image prominent.

$$\frac{1}{8}\begin{array}{|c|c|c|} \hline -1 & -1 & -1 \\ \hline -1 & 16 & -1 \\ \hline -1 & -1 & -1 \\ \hline \end{array}$$

Figure 5.27. Image sharpening Filter

Two images shown in Figure 5.28 are of same class with resolution of 256x256, change in facial expressions in both the images is prominent in forehead and mouth part of images which has been highlighted in Figure 5.28 by applying high pass filter. These wide facial expression changes present in both the images may provide poor matching results. To evaluate the facial expression compensation through image resolution, resolution of both the images has been reduced to 128x128 using decimation process. To highlight the expression sharpening filter is applied and results are shown in Figure 5.28 and 5.29. Analysis of these images shown in Figure 5.28 reflects that as the image resolution of face image is reduced from 256x256 to 128x128, the facial changes in the surrounding area of image forehead and mouth are reduced and resultant images seem to be identical and their wide expressions variations are reduced at this reduced resolution as compared to original resolution.

This is one of the systematic methodology to approach to *min(Dim(f))* as both the images are almost expressionless and still having requisite features required for recognition. This change in image resolution provides good matching and improves recognition rate. This work shows that image resolution variation through image decimation up to a certain factor retains all the required information for recognition, provides varying expression compensation within class and at the same time provides dimension reduction. Second methodology to approach more closer to *min(Dim(f))* is by using accurate feature extraction through which most relevant facial features required for improved recognition results are extracted.

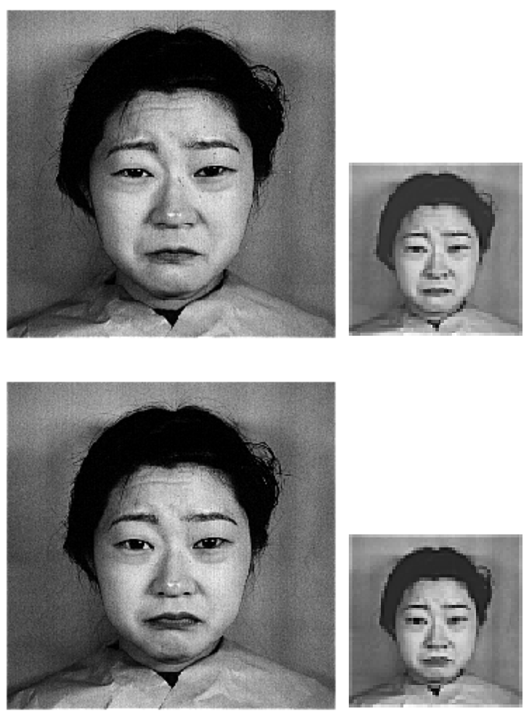


Figure 5.28. Two images of same class with changed facial expressions and corresponding images with reduced resolution.

Figure 5.29. Facial parts of images with varying expressions are highlighted

In case of DFT based face recognition method, low frequency feature extraction from real part of fourier spectrum is carried out through two different methods i.e square and circular method. As low frequency components fall around the center of the spectrum and carry maximum facial information required for recognition, these are extracted through these methods. Performing dimension minimization, in order to keep the dimensionality of an input pattern representation as small as possible, is advantageous for reasons such as minimizing computational cost and improving classification accuracy. However, a reduction in the number of features may directly lead to a loss in the discrimination power of a classifier and thereby yield lower accuracy in recognition. These schemes are used to approach closer to *min(Dim(f)),* to extract most suitable feature vector required for recognition with minimum data redundancy, maximum dimension reduction, best possible computation enhancement and improved recognition results.

To approach closer to *min(Dim(f))* the experiments have been carried out by varying dimension of square and circle feature vectors. Image resolution and dimension of feature vector with best results is taken as closest approach to *min(Dim(f)).*

In circle based feature selection method low frequency coefficients falling within a specified radius around the center of transformed image spectrum are taken as feature vector for recognition. Experiments with different dimensions of the radius have been carried out. Results shown in Figures 5.9 and 5.10 reveal that initially as we increase the radius or length of square recognition rate improves upto a certain radius. As the radius crosses this particular dimension the recognition rate start decreasing. This varying recognition rate with varying feature vector dimension is due to the fact that reaching at a particular dimension of feature vector against particular radius, it contains the minimum required features for better recognition. Below this particular radius, feature vector is lacking facial features required for recognition whereas above this radius recognition rate is decreased due to inclusion of redundant features in the feature vector which do not contribute towards recognition. Same is true for square selection method where the best square dimension is obtained against best recognition results.

Now once the best circle and square dimensions have been obtained, then these dimensions are kept constant and results are obtained by varying the image resolution. This reduction in image resolution is second step in refinement of required feature vector through further discarding the high frequency components to compensate varying facial expressions within same class. Results in Figures 5.10 and 5.12 show that image resolution has a certain effect on face recognition because as the image resolution is varied the recognition improves till a certain level where the best facial features with certain compensations are more suitable for recognition.

The most appropriate feature vector of facial image for recognition is obtained through double refinement process. First high frequency facial features are discarded through image decimation and secondly images are transformed through DFT and low frequency feature extraction is carried out by using square and circular feature extraction methods. Figure 5.5 reflects that only 121 features out of 10304 are used for recognition which is a considerable dimension reduction with improved computation speed.

DCT has many important properties that help significantly in image processing, especially in image compression. DCT itself does not compress the image, however, it can greatly help in achieving the desired compression or it can make the compression process considerably easier. One of these properties is energy compaction. Energy compaction means that the signal energy is concentrated or compacted on a few components (top left corner) while most other components are zero or are negligibly small. In DCT, lower frequencies are on the top left corner and they contain most of the information. The high frequencies don't have much of the information needed to reconstruct the image. Another important property of DCT is decorrelation. This type of frequency transform is real, orthogonal and separable. Algorithms for its computation have proved to be computationally efficient. Theses properties of DCT have made it the most suitable transform for image processing applications like compression and face recognition.

In DCT based face recognition method, the feature vector dimension is made closer to *min(Dim(f))* by reducing the image resolution as well as taking upper left corner coefficients of DCT spectrum as feature vector for recognition. Figure 5.17 reflects that 8x8 upper left corner coefficients of spectrum gives best FRR. At the same time it is established in Figure 5.17 that varying image resolution also provide improvement in FRR. Table 5.9 narrates that with reduced image resolution a considerable improvement in system training time and recognition time is achieved which makes the model computationally economical. In DCT based face recognition model, first face is exposed to image decimation process for resolution reduction where averaging (low pass) filter reduces the presence of high frequencies in the image. Then DCT with strong capability of energy compaction surrounds low frequency components in upper left corner of spectrum. Combination of these two methods provide better recognition results with improved model computation speed.

Wavelet are used to achieve multiresolution which is a powerful signal analysis tool and widely used in image compression and facial feature extraction applications. Images have locally varying statistics that result from different combinations of abrupt features like edges, of textured regions and of relatively low-contrast homogeneous regions. While such variability and spatial non stationarity defies any single statistical characterization, the multiresolution components are more easily handled. Wavelet transform can be performed for every scale and translation (resulting in continuous wavelet transform (CWT)), or only at multiples of scale and translation intervals (resulting in discrete wavelet transform (DWT)). An appropriate wavelet transform can result in robust representations with regard to lighting changes and is capable of capturing substantial facial features while keeping computational complexity low. From these considerations, in this research work DWT has been used to decompose face images into 4 subbands, denoted by LL, HL, LH, HH. The subband denoted by LL is approximately at half resolution of the original image. While the subbands HL and LH contain the changes of image or edges along vertical and horizontal directions, respectively. Subband HH contains the detail in the high frequency of the image and choose the lowest resolution subband coefficients for face representation. In DWT, the most prominent information in the signal appears in high amplitudes and the less prominent information appears in very low amplitudes. This high amplitude data is low frequency component of transformed spectrum with maximum information whereas high frequencies with finer details are with less amplitude. These low frequency components are used for face recognition purposes. Selection of suitable DWT family for recognition has been done by analyzing the different properties of DWT families as shown in Table 5.12. Later on different DWT of selected families are applied on face images and recognition is carried out and Symlet4 with best results as shown in Tables 5.10 and 5.11. Next issue is dimension reduction of face images with improved recognition rate and better computation speed. The complexity of a classifier grows rapidly with the number of dimensions of the pattern space and the computation speed is influenced accordingly. As the dimension of face images is reduced by increasing decimation down scale factor, the model training and recognition time is reduced. But at the same time it is important to keep the dimension reduction up to a limit where the most essential so-called discriminatory information which is conveyed by the extracted features is retained to improve the recognition results. A trade off between feature vector dimension reduction and best recognition results is obtained. First, image dimension drop through image resolution reduction, such a resolution is achieved where best recognition rate is achieved. Than DWT is applied to restrict the feature vector length to approximate coefficient (LL) part of transform spectrum. Computation speed of the system is further improved through feature optimization by defining a threshold value below which transformed values are made zero.

In this model of face recognition, the high frequency components are discarded twice, firstly in decimation process through averaging filter and secondly by applying DWT and considering approximate portion of spectrum for recognition. At the same time curse of dimensionality is addressed two fold: (1) through image decimation and (2) applying Symlet4 DWT on images with reduced resolution. With the combination of these two processes, improved FRR with improved training and recognition time has been recorded.

CHAPTER 6

DIMENSION REDUCTION ISSUES IN HEXAGONAL PIXELS

6.1 Introduction

The resolution reduction method used with square images in preceding chapters is now extended to hexagonal images. A new technique based on Diagonal grow and Butterfly structure methodology has been developed for sampling and indexing hexagonal structure in hexagonal image processing frame work. Proposed strategy offer less pixel redundancy as compared to existing techniques. Reduction in pixel redundancy varies according to size of square image.

Besides the conventional square grid for sampling and representing digital images hexagonal structure is an alternative pixel tessellation scheme. Sampling on a hexagonal lattice is a promising solution which has been proved to have higher efficiency and lesser aliasing. Hexagonal coordinate system rejects the common square grid upon which most images are mapped. Hexagonal coordinate systems are more significant as they bear a close resemblance to the layout of photo-receptors in the human eye retina. Research on the issue suggests that the simulation of at least some of the capabilities possessed by the human eye and the visual processing areas of the brain can be more easily executed on images that are laid out on a hexagonal lattice as compared to others. Another aspect which makes hexagonal structure much practical approach is the distance between neighboring points. In square lattice it depends on whether the neighbour is directly adjacent (in which case distance is 1) or diagonal (where distance is square root of 2) whereas in hexagonal system, all points are equidistant, at 1 unit. Due to this symmetry the hexagons are packed very close together and result in better picture quality and greater pixel density. Hexagonal structure provides better image processing transformations and operations like DFT, DCT and edge detection as compared to square image processing. There are four major considerations while processing image in hexagonal structure:

- Conversion from square to hexagonal structure
- Addressing and storing mechanism
- Image processing operations
- Image displaying mechanism

Hexagonal structure for image processing have many advantages like higher efficiency, lesser aliasing [118], reduced computational load, high circular symmetry, greater angular resolution and intelligent vision power. In spite of all these advantages, hexagonal grid has so far not been widely used in computer vision and graphics fields. The main problem that limits the use of hexagonal image structure is lack of hardware for capturing and displaying hexagonal-based images.

Designing a hardware to capture images from the real world directly onto a hexagonal lattice is highly specialist job, and not generally available for use. Hexagonal points are generated by mapping sampled points across an appropriate sampling lattice. Hexagonal lattices are usually denser than square equivalents which necessitates the extrapolation of additional points.

6.2 Hexagonal Structure Sampling and Addressing

Conversion from a square lattice to a hexagonal lattice is known as image re-sampling [119]. A number of efficient schemes for conversion of a standard square image into a hexagonal lattice have been developed [120 , 121, 122, 123]. A basic sampling lattice can be generated using the basis vectors (B = {b1, b2}) where b1={(1,0),b2= $(1/2, \frac{\sqrt{3}}{2})\}$. Most recent approaches are [124, 125] where least squares approximation of splines has been proposed for resampling square images onto a hexagonal lattice. This is a computationally intensive and better method.

The specific storage and addressing scheme can often drastically affect the performance of system. There are several ways to store and address a hexagonally-represented image. A simple approach is to use a 3-element coordinate system [126,127] where the 3 axes are aligned with the three axes of symmetry present in a hexagon within the lattice.

Skewed (x or y) axis [128, 129] is one more methodology where two axes are not orthogonal to each other. Scheridian proposed another addressing strategy [130] which is called single indexing system.

In most cases image processing transformations and operations (like distance measures, edge detection, convolution, Fast Fourier Transform, Discrete Fourier Transform) in hexagonal domain are quite different from square domain as both have different indexing, storage and neighborhood definitions [131, 132, 133]. In hexagonal structure points are not aligned in two orthogonal directions, so points in hexagonal grid do not lend themselves to be addressed by integer Cartesian coordinates. Different schemes have been presented to address the hex pixels, few are discussed below.

6.2.1 Skewed Axis

The simplest technique to address hex pixels is to use pair of skewed axis which is aligned along the axis of rotational symmetry of hexagon. It has two possibilities as shown in Figure 6.1 where axes are 120^{o} apart or 60^{o}. More variation is possible by rotating the axis in such a way that one of them is vertical. The examples of skewed axis can be found in [134,135,136].

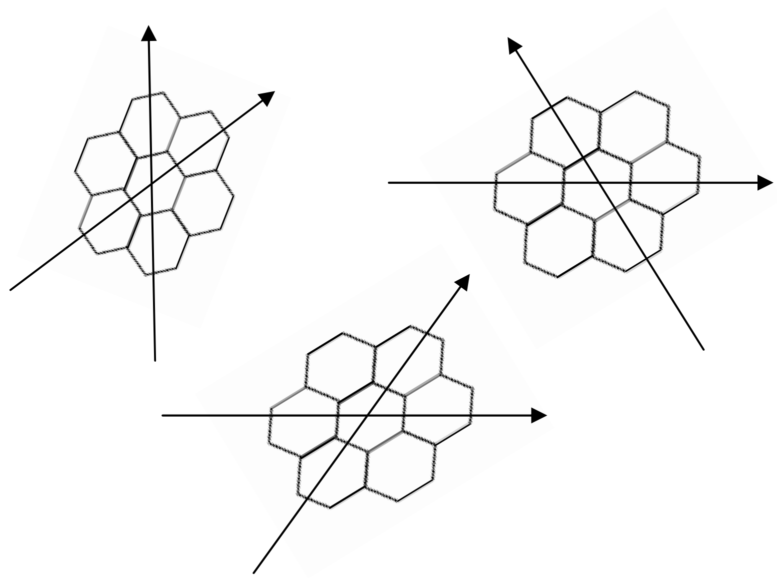


Figure.6.1 Few configurations of Skewed Axis

6.2.2 Three Axis of Symmetry

In this approach three axes (shown in Figure 6.2) of symmetry are used instead of two, where third axis is linear combination of other two axes due to which this indexing scheme suffers redundancy [137]. This addressing methodology uses a tuple of coordinates (a,b,c) which crossponds to the distance from the lines x=0, y=0 and z=0 respectively and obey the rule:

$$a+b+c=0 \tag{6.1}$$

The distance between any two neighbouring points in this scheme is 1. This coordinate scheme leads to more cumbersome data structure than using skewed axes and can lead to increased processing time, especially for non symmetric operations.

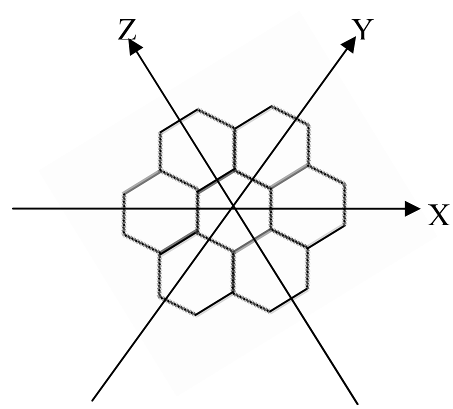


Figure.6.2. Three skewed Axis

6.2.3 Half Pixel Shift Approach

An approximation to hexagonal lattice is a brick wall where the pixel in alternative row are shifted by half a pixel to simulate the hexagonal lattice. Different methodologies [138, 139] have been adopted to implement this configration. Overington [140] treated entire hexagonal array as if it is a rectangular array and cartesian coordinates are directly employed to address all points. He moved either odd or even rows by half pixel.

6.2.4 Pyramidal Address Approach

This coordinate scheme uses a tuple (i, j, k) where last k is the level of pyramid and remaining two coordinates are same as in skewed axes methodology. Gibson and Tanimoto [141] also investigated hexagonal pyramid structure for the purpose of decomposition of hexagonal lattices and modeling.

6.2.5 Single Indexing System

One-dimensional addressing system proposed by Sheridan [142] also called spiral addressing strategy. In this approach hex address grows from the centre of image in powers of seven along a spiral like curve. Two mathematical operations, spiral addition and spiral multiplication [143] correspond to image translation and image rotation respectively.

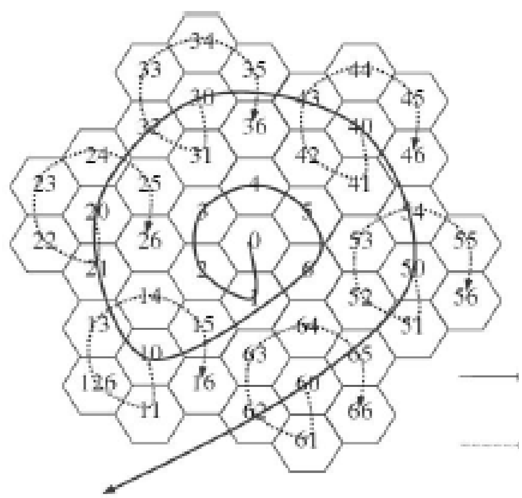


Figure.6.3 Single indexing Scheme

6.2.6 Middleton and Sivaswamy Approach

Generalized Balanced Ternary system based addressing approach was also discussed by Middleton and Sivaswamy [144]. Figure 6.4 shows the addressing scheme and use of ternary tree for finding the next pixel. This one dimensional addressing scheme leads to an efficient storage system and the placement of the origin at the centre of the image simplifies geometric transformations of a given image.

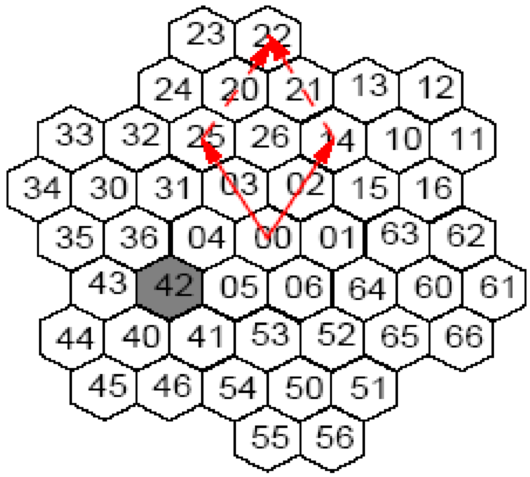
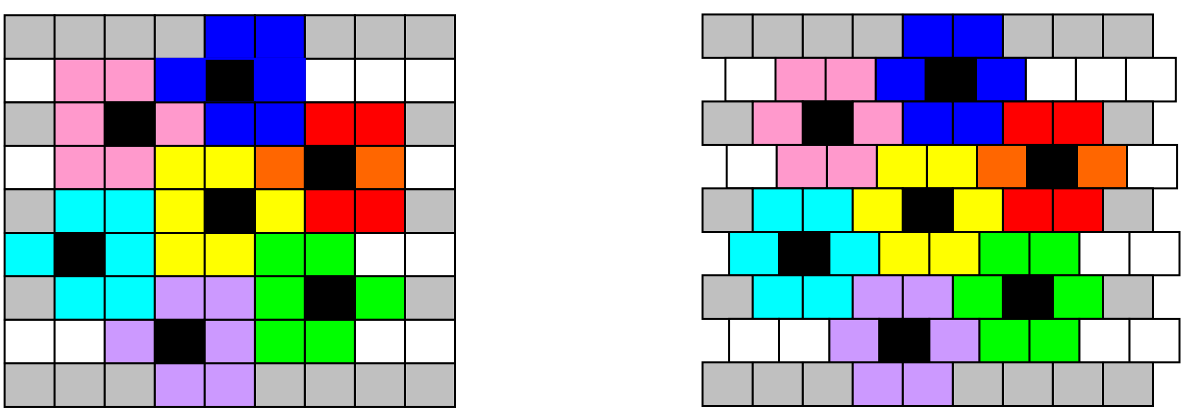


Figure 6.4. Middelton indexing Scheme

6.3 Proposed Hexagonal Processing Methodology

In proposed hex sampling and indexing methodology the half shift pixel strategy is adopted where each alternative row is shifted by half pixel. This half shift corresponds to a hex shape structure with one pixel in the center and six neighboring pixels as shown in Figure 6.5(b).



a b

Figure 6.5. (a) Pixels in square grid (b) Hexagonal Structure through half pixel shift

Figure 6.5(a) reflects the square pixels which contribute in the formation of hex structure. It is evident from Figure 6.5(a) that two types of patterns against hexagonal structure can be observed, one is right arm pattern and the second is left arm. Right and left arm patterns generated from a 3x3 square pixel, its half pixel shift and its corresponding hex pixels are reflected in Figure 6.6(a) and 6.6(b).

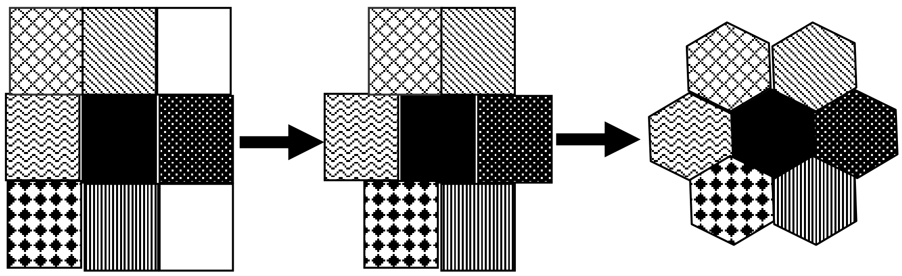


Figure 6.6. (a) Right arm hexagonal pixels with Corresponding Hexagons

Table.6.1. Conversion table for right arm pattern

Square Pixel Index	**Corresponding Hex Pixel Index**
(0,0)	(0,0)
(1,0)	(1,0)
(0,1)	$(0.5, \sqrt{\frac{3}{2}})$
(-1,1)	$(-0.5, \sqrt{\frac{3}{2}})$
(-1,0)	(-1,0)
(-1,-1)	$(-0.5, -\sqrt{\frac{3}{2}})$
(0,-1)	$(0.5, -\sqrt{\frac{3}{2}})$

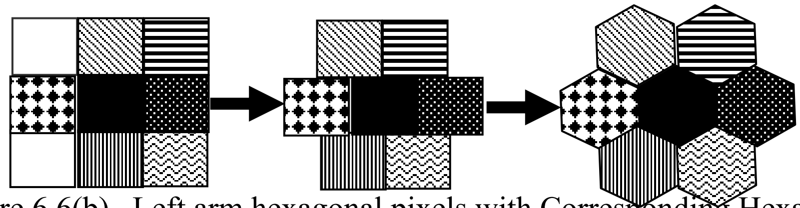


Figure 6.6(b). Left arm hexagonal pixels with Corresponding Hexagons

Table.6.2. Conversion table for left arm pattern

Square Pixel Index	Corresponding Hex Pixel Index
(0,0)	(0,0)
(1,0)	(1,0)
(1,1)	$(0.5, \sqrt{\frac{3}{2}})$
(0,1)	$(-0.5, \sqrt{\frac{3}{2}})$
(-1,0)	(-1,0)
(0,-1)	$(-0.5, -\sqrt{\frac{3}{2}})$
(1,-1)	$(0.5, -\sqrt{\frac{3}{2}})$

Figure 6.5 and 6.6 alongwith Table 6.1 and 6.2 illustrate the sampling of hex structure using proposed scheme. An image with dimensions 25x21 is shown in Figure 6.7 where it can be seen that combination of right and left arm patterns cover the whole square lattice in a specific fashion. The beauty of this sampling and indexing lies in direct linear conversion from square to hexagon where it takes directly the square pixel corresponding to hex pixels and place them in hex plane according to six neighboring arrangement from square lattice.

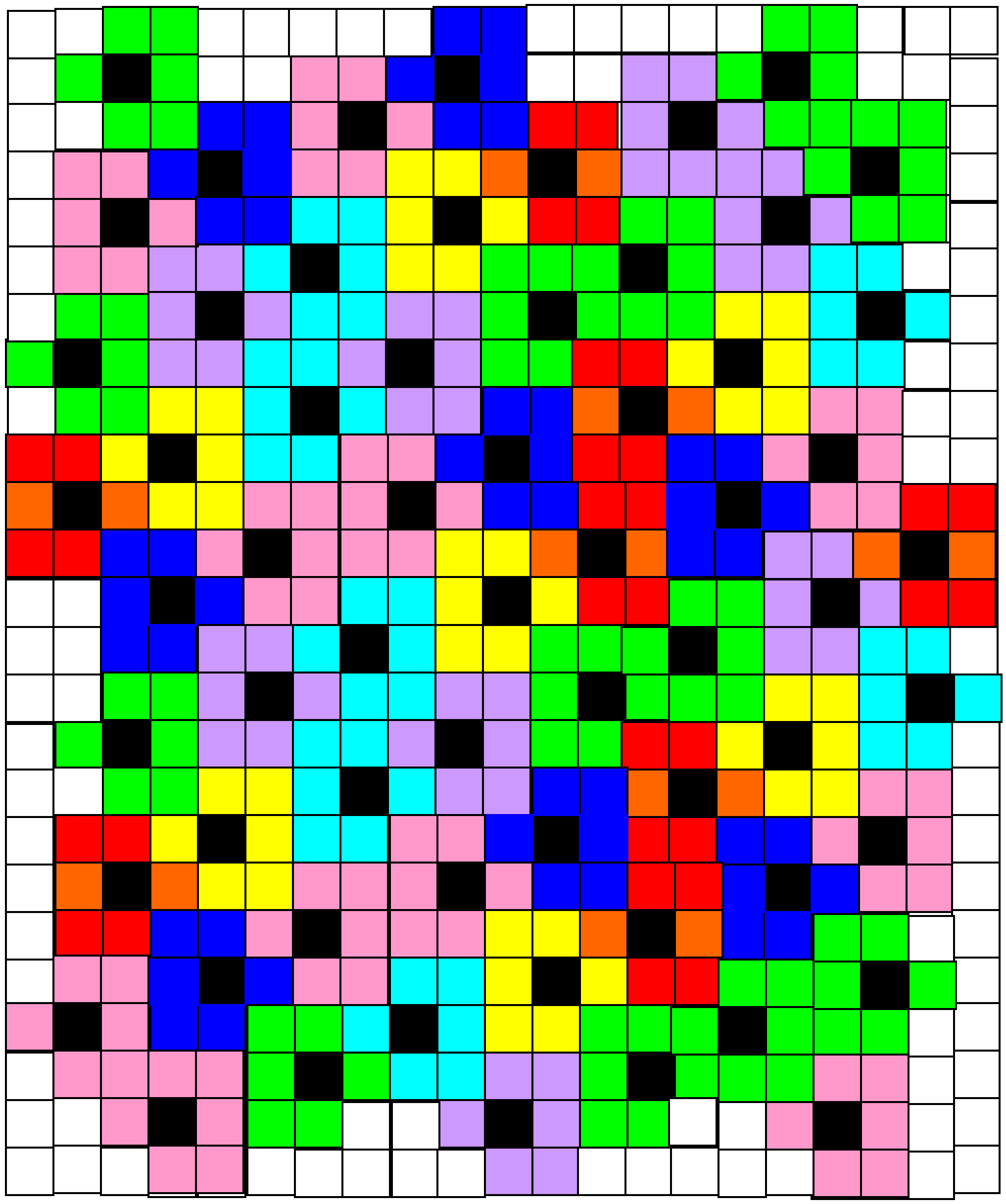


Figure 6.7. Selection of Hexagon Pixels from Square grid

6.3.1 Conversion Methodology

Hexagonal structure plane could be regularly sampled using integer multiples of a pair of basis vectors. In this approach first axis is taken parallel to the x-axis and a second rotated 120^{o} from the first, the basis vectors are:

$$B = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \ , \begin{pmatrix} \frac{-1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix} \right\} \tag{6.2}$$

These vectors provide a convenient coordinate system which can be used to further develop the tiling to produce an addressing scheme. In proposed scheme the two basic patterns right and left arm hexagonal patterns are obtained by shifting every alternative by half pixel. The coordinates for both patterns are:

$$P(r) = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\} \tag{6.3}$$

$$P(l) = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\} \tag{6.4}$$

where

$p(r)$ *is right arm pattern*

$p(l)$ *is left arm pattern*

Figure 6.7 shows that if centers of right arm and left arm patterns fall within 2nd & 3rd rows and 2nd & 3rd columns are calculated then sampling of complete square image to hexagon lattice can be obtained. This is due to the reason that centers of all subsequent right and left patterns are diagonal to these patterns. Let a square lattice be represented by I(x,y) then the hex centers will be

$$Hc_{r_{2(1)}} = I(a,b) \tag{6.5}$$

where

$a = 2 \text{ and } b = 3$

$Hc_{r_{2(1)}} = \text{First hex center of 2nd row as shown in Figure 6.13}$

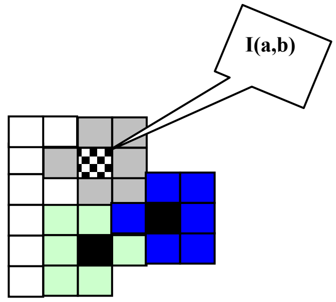


Figure 6.8. The first hex center (checker board) in second row

Every next hex center in same row will be after seven columns as shown in Figure 6.9 (pixel slanting lines pattern).

$$Hc_{r_{2(2)}} = I(d,e) \tag{6.6}$$

where

$$d = a \text{ and } e = b + 7$$

Whereas hex center in third row will be

$$Hc_{r_{3(1)}} = I(g,h) \tag{6.7}$$

where

$$g = 3 \text{ and } h = e - 2$$

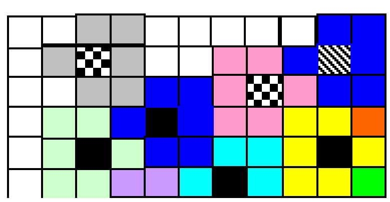


Figure 6.9. Hex center (checker board) in second & third row

Similarly the first four hex centers in 3rd and 2nd columns also lie in a specific pattern as shown in Figure 6.10.

$$Hc_{c_{3(1)}} = I(a,b)$$
$$Hc_{c_{3(2)}} = I(i,j) \tag{6.8}$$

where

$$i = a + 3 \text{ and } j = b$$
$$Hc_{c_{3(1)}} = \text{First hex center of 3rd column}$$

Similarly

$$Hc_{c_{2(1)}} = I(m,n)$$
$$m = i + 3 \text{ , } n = j - 1 \tag{6.9}$$
$$Hc_{c_{2(2)}} = I(m+3,n)$$

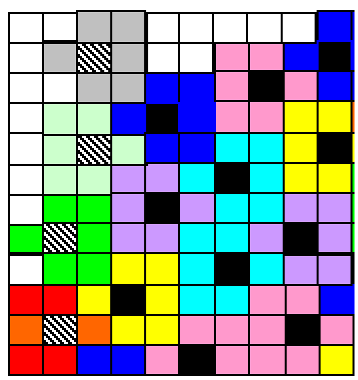


Figure 6.10. First four hex centers (checker board) in second and third column

This pattern of hex centers repeats itself after every four rows till the end of square lattice and similarly the growth of diagonals of these centers will let the complete image grow into hexagonal image. For storage two methods can be used, first is storage in one dimensional array for further processing and second is the direct mapping to the hex lattice.

6.4 Alternate Implementation of Proposed Indexing Scheme

Second method used to implement the proposed indexing scheme is quite different from first where a butterfly hex pixel pattern is created through a group of eight hexagons surrounded by six neighboring pixels which are combined in a specific way as shown in Figure 6.11

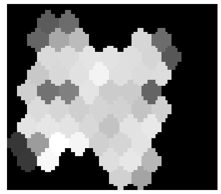


Figure 6.11. Butterfly pixel Pattern

In this method the centers of each hex of first butterfly pattern are obtained that lies in 13, 40, 25, 10, 37, 22, 7, 34 columns and 8, 13, 18, 23, 28, 33, 38, 43 rows. This pattern repeats itself in this specific order in every butterfly hex pixel and a butterfly hex pixel repeats itself after every 42 columns and 43 rows using intensity values with respect to square image. The algorithm developed for the conversion is as under:-

Algorithm Hexagonal structure conversion

Begin

1. Load input image
2. Initialize Hex image with respect to input image dimensions
3. Define boundary conditions with respect to input image dimensions
4. Define location of center of each hex pixels of butterfly hex pattern
5. Butterfly Hex conversion

Repeat till boundary conditions are met

- If row is even call left hex structure
- Else call right hex structure
- Take the intensity values and pass it to hex simulation function

6. repeat the butterfly hex structure pattern after 42 columns and 43 rows till boundary conditions

End

This is a simple conversion technique where square pixels are selected in a particular fashion and placed in butterfly structure in the form of hexagonal structure. Second technique involves formation of butterfly structure. Butterfly structure is combination of eight hexagons each with six neighbors as shown in Figure 6.9. Centers of each hexagon in butterfly structure are calculated which are:

$$\text{Butterfly Hexagonal Structure } (A1) = \begin{pmatrix} 8 \\ 13 \end{pmatrix} L, \begin{pmatrix} 13 \\ 40 \end{pmatrix} R, \begin{pmatrix} 18 \\ 25 \end{pmatrix} L, \begin{pmatrix} 23 \\ 10 \end{pmatrix} R, \begin{pmatrix} 28 \\ 37 \end{pmatrix} L, \begin{pmatrix} 33 \\ 22 \end{pmatrix} R, \begin{pmatrix} 38 \\ 7 \end{pmatrix} L, \begin{pmatrix} 43 \\ 34 \end{pmatrix} R \qquad (6.10)$$

Where R and L represents the right and left arm pattern

Butterfly hex repeats itself after every 42 columns and the pattern of repetition is:-

$$\text{Butterfly Hex repetition in column}(\text{Ac}_{\text{n}}) = A1 + \begin{pmatrix} 8 \\ 55 \end{pmatrix}, A1 + \begin{pmatrix} 8 \\ 97 \end{pmatrix}, A1 + \begin{pmatrix} 8 \\ 139 \end{pmatrix}, \ldots\ldots EOI \qquad (6.11)$$

where

Ac is repetition of Butterfly hex pixel in column which is after every 42 Columns

n is order of repetition

EOI is end of image

This repetition will continue till boundary conditions are met which defined are according to dimensions of input image. Similarly butterfly hex structure will grow in rows with specific pattern where difference of 7 columns in alternative repetition is due to shape and placement of this structure. The pattern will follow the rule.

$$\text{Butterfly Hex repetition in row}(Ar_n) = A1 + \begin{pmatrix} 51 \\ 13 \end{pmatrix}, A1 + \begin{pmatrix} 94 \\ 20 \end{pmatrix}, A1 + \begin{pmatrix} 137 \\ 13 \end{pmatrix}, A1 + \begin{pmatrix} 180 \\ 20 \end{pmatrix} \ldots\ldots EOI \quad (6.12)$$

where

Ar is repetition of Butterfly hex pixel in rows which is after every 42 rows

n is order of repetition

Storage of Hexagonal pixels

Once the hexagon center is obtained using above mentioned method, its six neighbours as shown in Figure.6.12 are obtained using following conversion equations shown in Table 6.3 and 6.4 for right arm and left arm patterns.

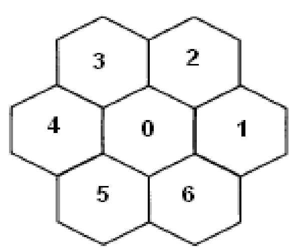


Figure.6.12. Hexagon pixel distribution

If storage is in an array then first element of array will be the center of the hexagon and next will be its six neighboring and similarly the array will grow till the end of the hex conversion. In case of storage in matrix form then the hex pixels will fall into matrix according to their position in the square lattice.

Table 6.3 Right arm Sampling

Table 6.4 Left arm Sampling

Right pattern Sampling
$C_{hex} = 0$
$C_{hex} + 1=1$
$C_{hex} – C_T = 2$
$C_{hex} - C_T – 1 = 3$
$(C_{hex} –1) = 4$
$C_{hex} + C_T-1 = 5$
$C_{hex} + C_T =6$

Left pattern Sampling
$C_{hex} = 0$
$C_{hex} + 1 = 1$
$C_{hex} - C_T + 1 = 2$
$C_{hex} - C_T = 3$
$C_{hex} - 1 = 4$
$C_{hex} + C_T = 5$

where

C_{hex} is center of hexagon, C_T is total column in the square grid

6.6 Simulation Results

Once the pixels have been sampled on hexagonal lattice through proposed hexagonal sampling methods, they are simulated on square screen using combination of square pixel in the form of hexagon. A hexagonal pixel constructed in this manner [14] is pleasing to eyes as it takes into account the oblique effect in human eye [145]. One example of construction of such hexagonal structure is shown in Figure 6.8 where a 7x6 square lattice is converted into single hexagonal pixel by discarding few square pixels. The combination of center hex pixel along with its six neighboring pixels is also shown in Figure 6.13.

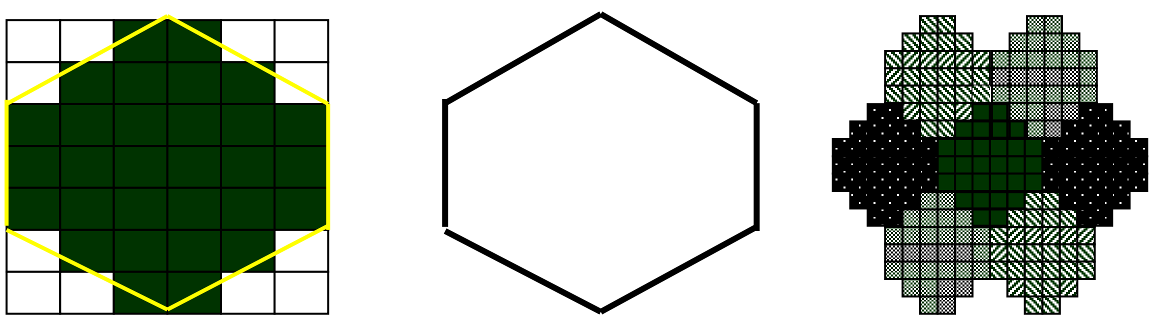


Figure.6.13. Hexagonal simulation through 7x6 square pixels (left) Hexagonal structure (center) Combination of center hex and its six neighboring hex pixels (right)

The face image of ORL database with dimension 112x92 was sampled into hexagonal lattice through proposed technique and simulation was carried out using hexagonal pixel generation shown in Figure 6.9. Results of this simulation are shown in Figure 6.14.

Figure.6.14. Face image in square lattice (left) converted in to hexagonal lattice (right)

6.7 Pixel Redundancy Comparison

In proposed methodology of diagonal grow hex sampling, the basic pattern of dimension 15x9 shown in Figure 6.15 repeats itself after nine columns and fifteen rows.

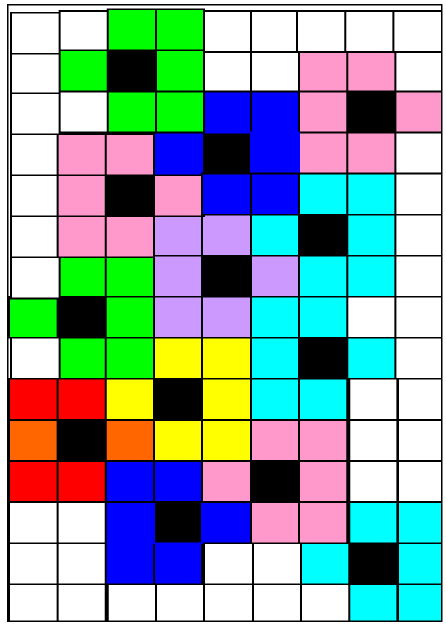


Figure 6.15. Basic pattern of diagonal grow scheme

Total numbers of pixels in the image of 15x9 are 44. Once this pattern repeats itself in columns the redundancy of pixels with each repetition will increase with multiple of 7 and 9 pixels at top and bottom respectively. Similarly when pattern repeat itself in rows the redundancy of pixels on right and left will be 14 and 11 respectively.

For comparison the image sampling from square to hex is also performed by adopting the indexing methodology of Lee Middleton [14]. This methodology starts the indexing from center and grow spirally out ward in layer approach of 7^{λ} where λ is number of layer. The conversion result of this approach is shown in Figure 6.16

Figure 6.16. Hex conversion using Middelton indexing approach.

In Middelton approach pixel redundancy is much more as compared to proposed approach.as:-

$$\text{Middelton Approach hex image grow with relation to level of layer} = 7^{\lambda}$$
$$\text{Where } \lambda \text{ is level of layer}$$

Let an image with dimensions of 15x9 (total 135 pixels) are required to sampled by using the Middelton layered approach if two layers are used then total pixels $7^2 = 49$ can be sampled which are insufficient to sample 135 pixels. So minimum option available will be to use third layer with total pixels of = 7^3=343. Now 343 pixels will be used to sample 135 pixels where 208 pixels will be redundant. By contrast proposed scheme require 179 pixels with only 44 redundant pixels. Table 6.5 gives the comparison of proposed technique and Middelton approach.

Table 6.5 Comparison of Redundancy in Proposed and Middelton Technique

Size of Image (Pixels)	Middelton Approach Total Pixels Required	Redundancy (%)	Proposed Approach Total Pixels Required (approximately)	Redundancy (%)
15x9 (135)	$\lambda^3 = 7^3 = 343$	60.6 %	169	26%
30x18 (540)	$\lambda^4 = 7^4 = 2401$	77.5%	625	13.6%
112x92 (10304)	$\lambda^5 = 7^5 = 16807$	38.9%	10619	2.9%

CHAPTER 7

CONCLUSIONS

7.1 In this research work challenges of automatic scale normalization, facial tilt present in the face image, dimension reduction and choice of minimum feature vector with best possible face recognition rate are addressed. An algorithm by using outer curvature of face is developed to retain only the facial part from the image to reduce the dimension with out missing any useful facial information required for recognition. In next phase of preprocessing the facial tilt present in the face image is removed by first locating the position of eyes in the image and then calculating the angle and slope present in both the eyes. This gradient is removed through reverse rotation.

7.2 In dimension reduction process, image resolution is reduced through image decimation to attain a resolution that minimizes the adverse effects of varying expressions and provided improved recognition result with reduced computational complexity. To achieve this, image is convolved with a low pass filter. Thereafter, the image decimation process is carried out to reduce the image resolution. To investigate the effects of reducing image resolution on recognition rate template matching technique is used. A matrix is obtained during training of system each column of which represents a single image with reduced resolution. Euclidean distance measure is used as matching criteria. Four different datasets (ORL, YALE, CEME_ NUST and FERET) are used to evaluate the effects of varying resolution on recognition. Experiments were carried out by varying the image resolution, an analysis of recognition rate, training time and recognition time against image resolution is performed and best recognition rate of 97.2% on ORL dataset on image resolution of 56 x 46 is achieved. CEME_ NUST color dataset is generated during this research work is also used to evaluate the effect of varying image resolution on recognition rate.

7.3 The effects of dimension reduction on recognition rate has also been explored using PCA. An improved recognition rate of 87% on ORL dataset is achieved. Moreover a hybrid technique in the form of Sub-Holistic PCA is developed where input decimated image is split in to four subparts and recognition process is carried on all sub faces individually. If four of the face spaces produce the same subject as the best match then the result of the four face spaces is considered to be the best match. Otherwise, the five resultant subjects/classes are used as training set for the DCT based classification which resulted in improved recognition rate. This methodology further improved the recognition up to 97% at the cost of computational complexity.

7.4. Novel preprocessing methods and facial expressions compensation through varying image resolution was used with DFT, DCT and DWT based face recognition methods. Moreover feature extraction through innovative methods is employed to achieve better recognition results. In DFT based face recognition two distinct methods (square and circular) have been employed to extract feature vector from DFT spectrum of face image. Extracted coefficient feature vector represent the facial features adequately that are required for better recognition. Moreover these methods results in reduced dimension vector. Computational complexity is reduced in consequence. Experiments have been carried out by varying the image resolution and keeping the feature vector dimension constant as well as by keeping the image resolution constant and varying the dimension of facial feature vector. YALE dataset at resolution of 56x 46 and with feature vector dimension of six provided 100% recognition rate.

Results with DCT based approach were obtained keeping the image resolution constant while varying the dimension of the feature vector and then varying the image resolution and fixing the dimension of feature vector. Outcome of these experiments revealed that there is little improvement in recognition rate with images at a specific resolution but a considerable reduction in computation time.

Discrete wavelet transform was employed for face recognition, where an image was represented in terms of translations and dilations of a scaling function and a wavelet functions. Different families of wavelet are used and symlet4 is observed the best choice. Feature vector optimization was also carried out to further reduce the dimension of feature vector while preserving the information required for recognition. The experiments were carried out using the images from five different datasets (ORL, YALE, NUST_EME, FERET and) with varying resolution and found 98.7% recognition rate at image resolution of 56 x 46.

7.5 The main philosophy behind all this work has been to evaluate the contribution of reduction in image resolution (through decimation) towards attainment of improved recognition rate. At the same time to achieve a feature vector which bears minimum feature vector dimension that reduce the computational complexity and results in faster processing. This idea of improved results with faster computation speed is materialized through:-

- Automatic image scale normalization which reduces the redundant data in the image retaining only facial portion of it.
- Achieving such image resolution using image decimation which gives maximum compensation to facial expression changes within class with out loosing any useful information.
- Applying unique feature extraction methodologies without any concession to useful information required for recognition.
- Feature vector optimization to further reduce any redundancy.

7.6 The resolution reduction issue of square pixel images is then extended to hexagonal pixel images. A novel hex structure sampling and indexing developed where the redundancy in existing hex structure conversion and indexing is reduced. The proposed hex sampling and indexing scheme is implemented through two different methodologies. First diagonal grow approach is used and in second methodology butterfly hex structure is constructed which repeats itself with a certain interval and complete image is converted in to hexagonal image. This linear hex technique minimizes the presence of redundant pixel in the simulation.

References

[1] S. Watanabe, Pattern Recognition: Human and Mechanical. Technical repot, New York: Wiley, 1985.

[2] R. Picard, Affective Computing. Technical report MIT Press, 1997.

[3] U. Grenander, General Pattern Theory. Tehnical report, Oxford Univ. Press, 1993.

[4] R. Bajcsy and S. Kovacic, ªMultiresolution Elastic Matching,Computer Vision Graphics Image Processing, vol. 46, pp. 1-21, 1989.

[5] L. Devroye, L. Gyorfi, and G. Lugosi, A Probabilistic Theory of Pattern Recognition. Berlin: Springer-Verlag, 1996.

[6] R.O. Duda and P.E. Hart, Pattern Classification and Scene Analysis,New York: John Wiley & Sons, 1973.

[7] K.S. Fu, Syntactic Pattern Recognition and Applications. Englewood Cliffs, N.J. Prentice-Hall, 1982.

[8] E. Oja, Subspace Methods of Pattern Recognition, Letchworth, Hertfordshire, England: Research Studies Press, 1983.

[9] J. L. Wayman, "Fundamentals of Biometric Authentication Technologies", International Journal of Image and Graphics, Vol. 1, No. 1, pp. 93-113, 2001.

[10] CNN World News, "Schiphol Backs Eye Scan Security", March 27 2002. Available at http://www.cnn.com/2002/WORLD/europe/03/27/schiphol.security/.

[11] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "FVC2002: Fingerprint Verification Competition", Proc. International Conference on Pattern Recognition (ICPR), pp. 744-747, Quebec City, Canada, August 2002.

[12] P.J. Philips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone, "FRVT 2002: Overview and Summary", available from http://www.frvt.org/FRVT2002/documents.htm.

[13] M. Golfarelli, D. Maio and D. Maltoni, "On The Error-Reject tradeoff in Biometric Verification Systems", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.19, No.7, pp. 786-796, July 1997.

[14] M. Golfarelli, D. Maio and D. Maltoni, “On The Error-Reject tradeoff in Biometric Verification Systems”, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.19, No.7, pp. 786-796, July 1997.

[15] W. R. Harrison, Suspect Documents, Their Scientific Examination, Nelson-Hall Publishers, 1981.

[16] A. Eriksson and P. Wretling, “How Flexible is the Human Voice? A Case Study of Mimicry”, Proc. of the European Conference on Speech Technology, pp. 1043-1046, Rhodes, 1997.

[17] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, “Impact of Artificial Gummy Fingers on Fingerprint Systems”, Proc. SPIE, Vol. 4677, pp. 275-289, San Jose, USA, Feb 2002.

[18] A. K. Jain, A. Ross and S. Prabhakar,”An Introduction to Biometric Recognition” IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, Vol. 14, No. 1, January 2004.

[19] John D. Woodward, Jr., Christopher Horn,Julius Gatune, and Aryn Thomas,”Biometrics A look at facial recognition” A documented briefing prepared for Virginia State Crime Commission.

[20] Technical Report, “Biometrics: A look at Facial Recognition”, By John D. woodward, et al prepared for Virginia state Crime commission 2003.

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

[22] K.-C. Chung, S. C. Kee, and S. R. Kim, “Face recognition using principal component analysis of Gabor filter responses,” in Proc. International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFGRTS ’99), pp. 53–57, Corfu, Greece, September 1999.

[23] B. Moghaddam, T. Jebara, and A. Pentland, “Bayesian face recognition,” Pattern Recognition, vol. 33, no. 11, pp. 1771– 1782, 2000.

[24] X. Mu, M. Artiklar, M. H. Hassoun, and P. Watta, “Training algorithms for robust face recognition using a template matching approach,” in Proc. International Joint

Conference on Neural Networks (IJCNN '01), vol. 4, Washington, DC, USA, July 2001.

[25] K. Hotta, T. Mishima, T. Kurita, and S. Umeyama, "Face matching through information theoretical attention pointsand its applications to face detection and classification," in Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG '00), pp. 34–39, Grenoble, France, March 2000.

[26] R.-L. Hsu and A. K. Jain, "Face modeling for recognition," in Proc. International Conference on Image Processing (ICIP '01), vol. 2, pp. 693–696, Thessaloniki, Greece, October 2001

[27] J. Zhu, B. Liu, and S. C. Schwartz, "General illumination correction and Its application to face normalization," in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '03), vol. 3, pp. 133–136, Hong Kong, China, April 2003.

[28] G. J. Edwards, C. J. Taylor, and T. F. Cootes, "Interpreting face images using active appearance models," in Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition (FG '98), pp. 300–305, Nara, Japan, April 1998.

[29] Craw, I., Tock, D. & Bennett, A. (1992), Finding Face Features, in 'European Conference on Computer Vision', pp. 92–96.

[30] R. Brunelli and T.Poggio. "Face recognition: features vs templates". IEEE transactions in Pattern Analysis. March. Intell. 19 (1997) pages 675-676.

[31] Matthew Turk and Alex Pentland, "Eigenfaces for Face Recognition". Journal of Cognitive Neuroscience, Vol3, No.1, page 71-86. 1991

[32] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces". IEEE transactions on pattern analysis and machine intelligence, Vol 12, No.1, Jan1990.

[33] Aliex M. Martinez and Avinash C. Kak, "PCA versus LDA". IEEE transactions on Pattern Analysis and Machine Intelligence, Vol 23, No2. page 228-233, 2001.

[34] Jie Yang, Hua Yu and William Kunz, "An efficient LDA Algorithm for Face Recognition". Carnegie Mellon University.

[35] Marian Stewart Bartlett, Javier R. Movellan and Terrence J. Sejnowski, “Face Recognition by Independent Component Analysis”. IEEE Transactions on Neural Networks, Vol.13, No 6.November 2002.

[36] J. F. Cardoso, “Infomax and Maximum Likelihood for Source Separation”. IEEE Letters on Signal Processing, Vol 4. pages 112-114. 1997.

[37] Daniel D. Lee and H. Sebastian Seung, “Learning the parts of objects by non-negative matrix factorization”. Nature Vol 401, 21 October 1999.

[38] Meng Joo Er,Shiqian Wu, Juwei Lu, Hock Lye Toh, ”Face Recognition With Radial Basis Function Neural Networks”. IEEE Transactions on Neural Networks, Vol 13, No.3, May 2002.

[39] Vapnik, V. Statistical Learning Theory, John Wiley & Sons, New York, United States(1998).

[40] Guo, G., Li, S. & Chan, K. (2000), Face recognition by Support Vector Machines, in ‘Proceedings of Fourth IEEE Conference Automatic Face and Gesture Recognition’, Grenoble, France, pp. 196–201.

[41] Xi, D., Podolak, I.&Lee, S. (2002), Facial component extraction and face recognition with support vector machines, in ‘Proceedings of Fifth IEEE Conference Automatic Face and Gesture Recognition’, Washington DC, United States, pp. 76–81.

[42] Tefas, A., Kotropoulos, C. & Pitas, I. (2001), Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication, in ‘IEEE Transactions on Pattern Analysis and Machine Intelligence’,Vol. 23 (7), Grenoble, France, pp. 735–746.

[43] Colmenarez, A. & Huang, T. (1998), Face Detection and Recognition, in H. Wechsler, P. J. Phillips, V. Bruce, F. F. Souli´e & T. S. Huang, eds, ‘Face Recognition: From Theory to Applications’, NATO ASI Series F, Springer-Verlag.

[44] A. Averbuch, D. Lazar, and M. Israeli, “Image Compression Using Wavelet Transform and Multiresolution Decomposition” IEEE Trans. Image Processing, vol. 5, no. 1, Jan. 1996

[45] R. Fotlyniewiez, “Automatic face recognition via wavelets and mathematical morphology”, Proceedings of ICPR 1996.

[48] W.S.Lee, H.J.Lee, J.H.Chung,.” Wavelet-based FLD for face recognition”, Proceedings of the 43rd IEEE Midwest Symposium on Circuits and Systems, 2000 Volume 2, 8-11 Aug. 2000 Page(s):734 - 737 vol.2

[47] Z.Ying; S.Schwartz,. M.Orchard, “Fast face detection using subspace discriminant wavelet features”, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2000.Volume 1, 13-15 June 2000

[48] K. Ma., and T. Xiaoou, “Discrete wavelet face graph matching”, Proceedings of International Conference on Image Processing, Volume 2, 7-10 Oct. 2001 Page(s):217 – 220.

[49] C. Liu; and H. Wechsler, “A Gabor feature classifier for face recognition”, Proceedings of Eighth IEEE International Conference on Computer Vision, ICCV 2001 Page(s):270 - 275 vol.2 7-14 July 2001.

[50] L.H.Yang.; T.D.Bui,and C.Y. Suen,”An application of nonlinear wavelet approximation to face recognition”. Proceedings of 16th International Conference on Pattern Recognition,Volume 2, 11-15 Page(s):48 - 51 Aug. 2002.

[51] L.Wiskott,J.M.Fellous,N.Kruger and C.V. Malsburg, ”Face recognition by elastic bunch graph matching”,IEEE transaction on pattern analysis and machine intelligence Vol 19 No 7, Jul 1997.

[52] R. Chellappa, C. L. Wilson, and S. Sirohey, “Human and machine recognition of faces: A survey,” Proc. IEEE, vol. 83, no. 5, pp. 705–740, 1995.

[53] A. Samal and P. A. Iyengar, “Automatic recognition and analysis of human faces and facial expression: A survey,” Pattern Recognition, vol. 25, no. 1, pp. 65–77, 1992.

[54] D. Valentin, H. Abdi, A. J. O’Toole, and G. W. Cottrell, “Connectionist models of face processing: A survey,” Pattern Recognition, vol. 27, no. 9, pp. 1209–1230, 1994.

[55] P.J. Phillips, H. Moon, et. al., “The FERET Evaluation Methodology for Face-Recognition Algorithms”, IEEE TPAMI, Vol.22, No.10, pp.1090-1104, 2000.

[56] H.F. Chen, P.N. Belhumeur, D.W. Jacobs, In search of illumination invariants, in: Proceedings of the Conference Computer Vision and Pattern Recognition, 2000, pp. I: 254–261.

[57] D.W. Jacobs, P.N. Belhumeur, R. Basri, Comparing images under variable illumination, in: Proceedings of the Conference Computer Vision and Pattern Recognition, 1998, pp. 610–614.

[58] M. Savvides, B.V. Kumar, Efficient design of advanced correlation filters for robust distortion-tolerant face recognition, in: Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS'03), 2003, pp. I: 254–261.

[59] W. Zhao, R. Chellappa, Illumination-insensitive face recognition using symmetric shape-from-shading, in: Proceedings of the Conference Computer Vision and Pattern Recognition, 2000, pp. IV: 297–301

[60] A. Shashua and T.R. Raviv, The quotient image: Class based re-rendering and recognition with varying illuminations, IEEE Trans. Pattern Anal. Mach. Intell. **23** (2001) (2), pp. 129–139.

[61] P.N. Belhumeur and D.J. Kriegman, What is the set of images of an object under all possible illumination conditions?, Int. J. Comput. Vision **28** (1998) (3), pp. 1–16.

[62] A.S. Georghiades, P.N. Belhumeur and D.J. Kriegman, From few to many: Illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. **23** (2001) (6), pp. 643–660.

[63] R. Ishiyama, S. Sakamoto, Geodesic illumination basis: compensating for illumination variations in any pose for face recognition, in: Proceedings of the International Conference on Pattern Recognition, 2002, IV: pp. 297–301.

[64] R. Basri, D. Jacobs, Lambertian reflectance and linear subspaces, in: Proceedings of the IEEE International Conference on Computer Vision, 2001, pp. II:384–390.

[65] L. Zhang, D. Samaras, Face recognition under variable lighting using harmonic image exemplars, in: Proceedings of the Conference Computer Vision and Pattern Recognition, 2003, pp. I:19–25.

[66] T.F. Cootes, G.J. Edwards and C.J. Taylor, Active appearance models, IEEE Trans. Pattern Anal. and Mach. Intell. **23** (2001) (6), pp. 681–685

[67] V. Blanz, S. Romdhami, T. Vetter, Face identification across different poses and illuminations with a 3D morphable model, in: Proceedings of the International Conference on Automatic Face and Gesture Recognition, 2002, pp. 202–207.

[68] D. Beymer, T. Poggio, Face recognition from one example view, in: Proceedings of the IEEE International Conference on Computer Vision, 1995, pp. 500–507.

[69] T.S. Jebara, 3D pose estimation and normalization, Ph.D. Thesis, McGill University, 1996.

[70] V. Blanz, and T. Vetter, "Face identification across different poses and illumination with a 3D morphable model," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, pp. 1063-1074, 2003.

[71] R. Gross, I. Matthews, and S. Baker, "Fisher light-fields for face recognition across pose and illumination," Proceedings of the 24th Symposium of the German Association for Pattern Recognition (DAGM), Zurich, Switzerland, 2002.

[72] M.A.O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," European Conference on Computer Vision, vol. 2350, pp. 447-460, Copenhagen, Denmark, May 2002.

[73] S. Zhou and R. Chellappa, "Rank constrained recognition under unknown illumination," IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003.

[74] S. Baker and T. Kanade, .Hallucinating Faces,. The fourth International Conference on Automatic face and Gesture Recognition ,2000.

[75] C. Liu, H.Y. Shum and C.S. Zhang, .A Two-Step Approach to Hallucinating Faces: Global Parametric Model and Local nonparametric Model,. In Proc. IEEE International Conference on Computer Vision and Pattern Recogntion, 2001.

[76] R. C. Gonzalez,. and R. E. Woods, "Digital Image Processing", *Addison–Wesely Publishing Company,* 1992

[77] G .Wyszecki. and S. W. Stiles, "Color Science : Concept and Methods, Quantitative Data and Formulas", *New York: Wiley*, 1982

[78] D.J.Beymer, “Face recognition under varying pose”, IEEE International conference on Computer Vision and Pattern Recognition ,pages756-761,1994.

[79] D.Reisfeld,N.Arad and Y. Yeshurn, “Normalization of face mages using few anchors”, International conference on pattern recognition 1994 Vol 1 pages 761-763

[80] J. Canny. “A Computational Approach to Edge Detection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 8, No. 6, Nov 1986.

[81] Kun Peng, Liming Chen, Su Ruan, and Georgy Kukharev, “A Robust Algorithm for Eye Detection on Gray Intensity Face without Spectacles” JCS&T Vol. 5 No. 3 October 2005.

[82] A.J.Candeas, and U.de.M.Braga, ”A Mathematical Morphology Approach to the Star / Galaxy Characterization ” *Journal of Brazilian Computer Society* vol3.no3April 1997.

[83] J.Serra, “Image analysis and mathematical morphology vol 1 Acdemic Press 1992”.

[84] Ferdinando Samaria, Andy Harter. Parameterisation of a Stochastic Model for Human Face Identification. Proceedings of 2nd IEEE Workshop on Applications of Computer Vision, Sarasota FL, December 1994

[85] P. N. Bellhumer, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, Special Issue on Face Recognition, 17(7):711--720, 1997.

[86] Color database obtained at *Image processing Lab at College of E&ME (NUST)*,Rawalpindi, Pakistan, Oct, 2004.

[87] P.J. Phillips, H. Wechsler, J. Huang, and P.J. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 1998, 16(5): 295-306.

[88] The CMU Pose, Illumination, and Expression Database T. Sim, S. Baker, and M. Bsat IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 12, December, 2003

[89] Jingdong Wang, Changshui Zhang, Heung-Yeung Shum. Face Image Resolution versus Face Recognition Performance Based on Two Global Methods. Asia Conference on Computer Vision (ACCV'2004), Jeju Island, Korea, Jan., 2004

[90] Pascalis, O., de Schonen, S., Morton, J., Deruelle, C. & Rabre-Grenet, M. (1995) *Inf. Behav. Dev.* 18, 79-85.

[91] Maxim A.Grudin, *"On internal representations in face recognition system"*. Pattern Recognition 33 (2000) page 1161-1177

[92] R. Brunelli and T.Poggio. *"Face recognition: features vs templates"*. IEEE transactions in Pattern Analysis. March. Intell. 19 (1997) pages 675-676

[93] Matthew Turk and Alex Pentland, *"Eigenfaces for Face Recognition"*. Journal of Cognitive Neuroscience, Vol3, No.1, page 71-86. 1991.

[94] H. Zhao; P. Chi Yuen and Kwok, J.T, "A novel incremental principal component analysis and its application for face recognition"(92.5) IEEE Transaction on Systems, Man and Cybernetics, Part BVolume 36, Issue 4, Aug. 2006 Page(s):873 – 886

[95] H.Y. Wang and X. J. Wu, "weighted pca space and its application in face recognition " Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005

[96] K.In.Kim, K.Jung and H.J.Kim,"Face recognition using kernel principal component Analysis",IEEE signal Processing letters vol9.No 2,February 2002.

[97] S.Shan, W. Gao, and Debin Zhao,"Face recognition based on face specific subspace",Wiley periodicals, Inc vol 13, 23-32. 2003.

[98] V.D. M. Nhat and S. Y. Lee, "Two-dimensional Weighted PCA algorithm for Face Recognition", Proceedings 2005 IEEE International Symposium on Computational Intelligence in Robotics and Automation June 27-30, 2005, Espoo, Finland.

[99] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America A*, vol. 4, no. 3, pp. 519–524, March 1987.

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

[101] Matthew A. Turk and Alex P. Pentland, "“Face recognition using eigenfaces.”," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Hawaii, June 1991, pp. 586–591.

[102] David Beymer, Amnon Shashua, and Tomaso Poggio, “Example based image analysis and synthesis”, A. I. Memo 1431, MIT, 1993.

[103] [7] H. Murase and S. K. Nayar, "“Visual learning and recognition of 3-D objects fromappearance”, " *International Journal of Computer Vision*, vol. 14, pp. 5–24, 1995.

[104] A. P. Pentland, B. Moghaddam, and T. Starner, "“Viewbased and modular eigenspaces for face recognition”," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994, pp. 84–91.

[105] L.D. Harmon, The recognition of faces, Sci. Am. 229 (1973) pp. 71-82.

[106] A.P. Ginsburg, “Visual information processing based on spatial filters constrained by biological data”, *AMRL Technical Report*, 1978.

[107] J. H. Lai!, P. C. Yuen. and G. C. Feng “Face recognition using holistic Fourier invariant features” *Pattern Recognition 34 (2001) 95}109.*

[108] A. Averbuch, D. Lazar, and M. Israeli, “Image Compression Using Wavelet Transform and Multiresolution Decomposition” *IEEE Trans. Image Processing,* vol. 5, no. 1, Jan. 1996

[109] R. Fotlyniewiez, “Automatic face recognition via wavelets and mathematical morphology”, *Proceedings of ICPR* 1996.

[20] W.S.Lee, H.J.Lee, J.H.Chung,.” Wavelet-based FLD for face recognition”, *Proceedings of the 43rd IEEE Midwest Symposium on Circuits and Systems*, 2000 Volume 2, 8-11 Aug. 2000 Page(s):734 - 737 vol.2

[111] Z.Ying; S.Schwartz,. M.Orchard, “Fast face detection using subspace discriminant wavelet features”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2000.Volume 1, 13-15 June 2000

[112] K. Ma., and T. Xiaoou, "Discrete wavelet face graph matching", *Proceedings of International Conference on Image Processing,* Volume 2, 7-10 Oct. 2001 Page(s):217 – 220.

[113] C. Liu; and H. Wechsler, "A Gabor feature classifier for face recognition", Proceedings of Eighth *IEEE International Conference on Computer Vision, ICCV 2001* Page(s):270 - 275 vol.2 7-14 July 2001.

[114] L.H.Yang.; T.D.Bui,and C.Y. Suen,"An application of nonlinear wavelet approximation to face recognition". *Proceedings of 16th International Conference on Pattern Recognition,*Volume 2, 11-15 Page(s):48 - 51 Aug. 2002.

[115] L.Wiskott,J.M.Fellous,N.Kruger and C.V. Malsburg, "Face recognition by elastic bunch graph matching",*IEEE transaction on pattern analysis and machine intelligence* Vol 19 No 7, Jul 1997.

[116] T. Chang and C.-C.J. Kuo, "Texture Analysis and Classification with Tree-Structured Wavelet Transform" *IEEE Trans. Image Processing*, vol. 2, no. 4,pp. 429-441, Oct. 1993.

[117] A. Averbuch, D. Lazar, and M. Israeli, "Image Compression Using Wavelet Transform and Multiresolution Decomposition" *IEEE Trans. Image Processing*, vol. 5, no. 1, Jan. 1996.

[118] Mersereau, R.M., The processing of Hexagonally Sampled Two-Dimensional Signals. Proceedings of the IEEE, 1979. pp. 930-949.

[119] G. Wolberg. Digital Image Warping. IEEE Computer Society Press, 1990.

[120] N.P.Hartman and S.L Tanimoto, "A Hexagonal Pyramid data structure for image processing," IEEE Transaction on systems,Man and Cybernetics, vol.SMC-14,pp.247-256,Mar-Apr 1984.

[121] A.B Watson and A.J. Ahumada, "A hexagonal orthogonal-oriented pyramid as a model of image representation in visual cortex," IEEE transactions on Biomedical Engineering, vol. BME-36,pp 97-106, Jan 1989.

[122] A.P. Fitz and R.J.Green, "Fingerprint classification using a Hexagonal fast fourier transform," Pattern Recognition, vol.29,no.10,pp.1587-1597,1996.

[123] I. Overington, Computer Vision: a unified biologically inspired approach. Elsevier Science Publishing Company, 1992.

[124] D. Van De Ville, R. Van de Walle, W. Philips, and I Lemahieu, "Image resamling between orthogonal and Hexagonal lattices," IEEE international conference on image processing ,pp.389-392, 2002.

[125] D. Van De Ville, T.Blu, M Unser, W. Philips, I Lemahieu and R. Van de Walle, "Hex-splines: A novel spline family for hexagonal lattices," IEEE Transaction on Image processing, vol.13,no 6, pp 758-772, 2004.

[126] Her, I., A Symmetrical Coordinate Frame on the Hexagonal Grid for Computer Graphics and Vision. ASME Journal of Mechanical Design, 1993. **115**(3): p. 447-449.

[127] Her, I. and C.T. Yuan, Resampling on a Pseudohexagonal Grid. CVGIP: Graphical Models and Image Processing, 1994. **56**(4): p. 336-347

[128] Luczak, E. and A. Rosenfeld, Distance on a Hexagonal Grid. IEEE Transactions on Computers, 1976. **25**(5): p. 532-533.

[129] Wuthrich, C.A. and P. Stucki, An algorithmic comparison between square- and hexagonal-based grids. CVGIP: Graphical Models and Image Processing, 1991. **53**(4): p. 324-339.

[130] Sheridan, P., Spiral Architecture for machine vision. 1996, University of Technology, Sydney.

[131] Middleton, L. and J. Sivaswamy, Edge detection in a hexagonal-image processing framework. Image and Vision Computing, 2001. **19**(14): p. 1071-1081.

[132] Middleton, L. and J. Sivaswamy, The FFT in a Hexagonalimage Processing Framework. Proceedings of Image and Vision Computing New Zealand, 2001: p. 231-236.

[133] Wu, H.-S., Hexagonal discrete cosine transform for image coding. Electronics Letters, 1991. **27**(9): p. 781-783.

[134] B.Mashes and W. Pearlman, "Image coding on a hexagonal pyramid with noise spectrum shaping,"Proceeding of International Society for Optical Engineering, vol.119,pp. 764-774,1989.

[135] S.C.M. Bell, F. C. Holroyd, and D.C. Mason, “A digital geometry for hexagonal pixels,” Image and vision computing, vol7,pp.194-204,1989.

[136] A.Nel, “Hexagonal image Processing,” in COMSIG, IEEE, 1989.

[137] I.Her, “Geometric Translation on the hexagonal grid,” IEEE transactions on image processing, vol4, pp 1213-1222,September 1995.

[138] A.P.Fitz and R.J. Green, “Fingerprint classification using a hexagonal FFT”, Pttern Recognition, vol.29,no 10, pp1587-1597, 1996.

[139] A.F. Lanie, S. Schuler, W.Huda, J.C. Honeyman, and B. SSteinbach, “Hexagonal wavelet processing of digital mamography ”, in proceedings of SPIE, vol.1898,pp.559-573,1993.

[140] I.Overington, Computer Vision : a unified biological inspired approach, Elsevier Science Publishing Company, 1992.

[141] L. Gibson and D. Lucas, “Spatial data processing using generalised balanced ternary,” in Proceeding of PRIP 82. IEEE computer Society conference on Pattern Recognition and Image Processing , pp.no.566-571, 1982.

[142] Sheridan, P., Spiral Architecture for machine vision. 1996, University of Technology, Sydney

[143] He, X., 2-D Object Recognition with Spiral Architecture. 1999, University of Technology, Sydney.

[144] Middleton, L. and J. Sivaswamy, Framework for practical hexagonal-image processing. Journal of Electronic Imaging, p. 104-114 2002.

[145] R.M.Gray, P.C.Cosman and K.L.Oehler, “Digital Images and Human Vision”, MIT press Cambridge, MA, 1993,cahpter 4,pp 35-52.