University of Windsor

Scholarship at UWindsor

Electronic Theses and Dissertations

Theses, Dissertations, and Major Papers

2008

A new approach to face recognition using Curvelet Transform

Tanaya Mandal
University of Windsor

A NEW APPROACH TO FACE RECOGNITION USING CURVELET TRANSFORM

by

Tanaya Mandal

A Thesis
Submitted to the Faculty of Graduate Studies
through Electrical and Computer Engineering
in Partial Fulfillment of the Requirements for
the Degree of Master of Applied Science at the
University of Windsor

Windsor, Ontario, Canada
2008

Your file Votre référence
ISBN: 978-0-494-42267-0
Our file Notre référence
ISBN: 978-0-494-42267-0

Author's Declaration of Originality

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University of Institution.

ABSTRACT

Multiresolution tools have been profusely employed in face recognition. Wavelet Transform is the best known among these multiresolution tools and is widely used for identification of human faces. Of late, following the success of wavelets a number of new multiresolution tools have been developed. Curvelet Transform is a recent addition to that list. It has better directional ability and effective curved edge representation capability. These two properties make curvelet transform a powerful weapon for extracting edge information from facial images. Our work aims at exploring the possibilities of curvelet transform for feature extraction from human faces in order to introduce a new alternative approach towards face recognition.

To

My motherland, India

My parents

&

My brother...

ACKNOWLEDGEMENTS

I express my deep and sincere gratitude to my advisor, Dr. Jonathan Wu for giving me this opportunity to work under his supervision; for his guidance, suggestion and support. I would like to thank my external reader Dr. Nihar Biswas for his priceless advices. I convey my sincere thanks to my program reader Dr. Huapeng Wu for his valuable comments. Also I would like to thank my friend Angshul for a conference publication and for all the technical discussions we had. Special thanks to my mother Dali Mandal, my father Sankar Mandal, my brother Jeet, my fiancé Anirban, my friends Payel, Shilpi and Sudeepta for their constant support and love. Finally, I take this opportunity to thank my group members, especially Rashid and Adeel, for their kind help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAM:	Active Appearance Model
ASM:	Active Shape Model
CNN:	Convolution Neural Network
CWT:	Continuous Wavelet Transform
DCT:	Discrete Cosine Transform
DLA:	Dynamic Link Architecture
EBGM:	Elastic Bunch Graph Matching
EP:	Evolution Pursuit
FDCT:	Fast Discrete Curvelet Transform
FFT:	Fast Fourier Transform
FLA:	Fisher Linear Analysis
FLD:	Fisher Linear Discriminant
HMM:	Hidden Markov Model
ICA:	Independent Component Analysis
JAFFE:	Japanese Female Facial Expression
KPCA:	Kernel Principal Component Analysis
KL:	Kerhunen-Loeve
kNN:	k Nearest Neighbour
LDA:	Linear Discriminant Analysis
LFA:	Linear Fisher Analysis
MLP:	Multi-Layer Perceptron

NFL:	Nearest Feature Line
NFP:	Nearest Feature Plane
NFS:	Nearest Feature Space
OAA:	One Against All
OAO:	One Against One
ORL:	Olivetti Research Labrotary
PCA:	Principal Component Analysis
PDBNN :	Probability Decision Based Neural Network
NN :	Neural Network
SOM :	Self Organizing Map
STFT :	Short Time Fourier Transform
SVM:	Support Vector Machine
USFFT :	Unequally Spaced Fast Fourier Transform

CHAPTER 1

INTRODUCTION

1.1 Introduction to Face Recognition

Face Recognition can be defined as a problem of identifying or verifying a person from still images or video sequences using a stored database of facial images. Usually, the input to a face recognition system is an unknown face; in identification problems, the system retrieves the identity of the input face from the database of known individuals; whereas in verification problems the system either accepts or rejects the claimed identity of the query face. Face recognition has been studied for more than 30 years now and it has emerged as one of the most successful applications of image analysis. Due to the growing interest in biometrics authentication, it has become a popular research area not only in the field of computer vision, but in neuroscience and psychophysics as well. Unlike other biometrics (fingerprints, iris etc.) face recognition is non-intrusive i.e. images can be taken, identified or verified even without the knowledge of the subject. Moreover, the data required for the system is human readable and can be collected easily with simple devices like camera. Face recognition has become a major issue especially in the past decade, due to its important real-world applications in areas like video surveillance, smart cards, database security, telecommunication, digital libraries and medical records.

Broadly, face recognition techniques can be divided into two categories, depending upon the type of images (still or video) being used in the recognition system. Our work is confined to recognition of faces from still images only. Usually, the available images are 2D intensity images of human faces, which are 3D objects. So this problem can be seen as a problem of recognising 3D objects from 2D images. A fully automatic face recognition system must perform the following three subtasks: face detection, feature extraction and recognition/identification. However, each of these subtasks itself is a separate area of research and concentrating on all of them simultaneously is difficult. Isolating the subtasks not only simplify our job but also enhance the assessment and advancement of the component techniques. So, instead of detecting faces from images, standard databases of faces have been used for the experiments; and our prime focus has been developing a new efficient feature extraction technique.

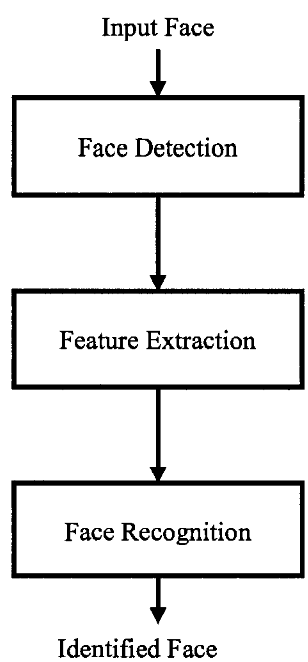

Figure 1.1: A Generic Face Recognition System

1.2 Challenges in Face Recognition

Faces are specific objects, that look similar from its most common appearance (frontal faces), but subtle features make them different. We, human beings, recognize faces with natural ease; but machine recognition of faces is a challenging job. The advantage of computer face recognition system is its capacity to handle large number of faces where human brain has limited memory i.e. capacity to remember faces. However, there are various problems associated with automatic face recognition; they are listed below:

- *Head pose:* Rotation or tilt of the head; even if the appearance is frontal, it affects the performance of the recognition system significantly.
- *Illumination change:* The direction of light illuminating the faces greatly affects the recognition accuracy. It has been noticed that illuminating a face image bottom up reduces the accuracy of the system.
- *Facial expression change:* From minor expression changes like smiling to extreme expression variations like shouting or crying or making grimaces, tend to affect the recognition rate of a system largely.
- *Aging:* Images taken at long interval or even at different days may seriously affect the correct recognition rate of a system.
- *Frontal vs. Profile:* Profile images can be difficult to recognize, when mostly frontal faces are available for matching.
- *Occlusion:* Even partial occlusion of faces due to objects or accessories like sunglass or scarf, makes identification a difficult task.
- *Hair style:* Change in hairstyle may affect the recognition accuracy as well.

Other issues with face recognition are: (i) it is not as accurate as other biometrics like fingerprints or iris recognition, (ii) optimum size of the facial images to be used is still an open issue to researchers, (iii) the problem with accurate feature localization. Though, static image based face recognition has attained a certain level, but still far away from the capability of human perception.

1.3 Objective

Feature extraction is a key step prior to recognition and an effective feature extraction method can greatly enhance the performance of any face recognition system both in terms of accuracy and speed. The features extracted from facial images can be local (lines, curves, edges etc.) or facial features (eyes, nose mouth etc.). This study aims at developing a novel face recognition technique from static images, employing a new multiresolution analysis tool called digital curvelet transform for feature extraction. The feature extraction method proposed in this thesis is a generic method which is based on capturing information about the local features like edges, curves etc.

1.4 Scope of this Work

A novel approach towards human face recognition based on a multiresolution analysis tool named digital curvelet transform is discussed in this thesis. This work includes the following studies:

- Face recognition using curvelet transform and k-NN classifier; comparison with wavelet based recognition method.

- Application of curvelet transform on bit quantized images and recognition using SVM classifier; comparison with results achieved using wavelet transform.
- Combination of curvelet features and PCA; comparison with existing techniques like traditional PCA and wavelet-based PCA.
- Combination of curvelet features with LDA; a comparative study with wavelet-based LDA method.
- A combined approach using both curvelet and wavelet transform, incorporating a PCA+LDA framework.
- Comparative study of proposed curvelet based methods with published work.

1.5 Thesis Outline

This thesis is organized into eight chapters. Chapter 1 provides a general introduction to face recognition, the challenges associated with this problem, the objectives and scope of the thesis. In Chapter 2 a detail review of previous work on face recognition can be found. Chapter 3 presents a theoretical description of wavelet transform and curvelet transform; as well as discusses why curvelet transform should be able to provide an effective means for feature extraction from face images. Chapter 4 aims at giving an overview of the classifiers that have been used for the identification task. Chapters 5, 6 and 7 explain the proposed methods and list the experimental results along with the comparative studies. Chapter 8 summarizes the contributions and indicates the scope of future work. Some additional information has been provided in the Appendices.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The earliest work on face recognition has been found to take place in 1950s in the field of psychology by Bruner and Tagiuri and in engineering by Bledsoe in 1960s. In 1970s, Kelly and Kanade started the research on automatic machine recognition of faces [1]. Since 1990s face recognition has become popular in different fields like computer vision, neuroscience, psychophysics etc. This field has grown fast and significantly in the past decade due to increased interest in commercial opportunities, availability of real-time hardware and importance in surveillance related applications.

Various techniques and algorithms have been developed by researchers. A comprehensive survey of all those techniques is out of the scope of this thesis. Interested reader can refer to the works of Zhao et al. [1] and Gross et al [2]. Since this work is restricted to the use of static images, the current chapter provides a detailed review on the still-image based face recognition research.

2.2 Feature Extraction Approaches

As illustrated in figure 1.1, feature extraction is a key step prior to face recognition. Extraction of proper features can greatly enhance the performance of any face recognition system. The features extracted from facial images are either local (lines, curves, edges etc.) or facial features (eyes, nose mouth etc.). Grossly, feature extraction techniques can be divided into three categories [1]:

- Generic Methods – based on edges, lines and curves.
- Feature-Template Based Methods – deals with detection of facial features like eyes, nose, mouth etc.
- Structural Matching Methods – a statistical method which takes geometrical constraints into consideration.

Early face recognition researchers focused on individual features, such as locating the position of eyes; later, structural matching methods were found to be more reliable. A successful example of such a statistical method is Active Appearance Model introduced by Cootes et al. [3] in 2001; this method is briefly discussed below.

2.2.1 Active Appearance Model [3]

A statistical model called 'Active Appearance Model' (AAM) [3] was suggested by Cootes et al. in 2001. This model is basically a combination of a shape model called Active Shape Model or ASM [4] and a model of appearance variation of shape free textures. A training set of 400 facial images, each manually labelled with 68 landmark points and 10,000 intensity values sampled from different facial images have been used for this work. Then, a shape model is constructed using each set of landmark points as a vector and applying PCA; a shape model consists of mean shape, orthogonal mapping matrix $\mathbf{P_s}$ and projection vector $\mathbf{b_s}$. Each sample image is warped to match its landmark points with mean shape. Thus, shape-free patches are formed to extract texture information. A texture model, consisting of mean texture, orthogonal mapping matrix $\mathbf{P_g}$ and projection vector $\mathbf{b_g}$, was constructed using PCA. To correlate between shape and texture variation, PCA is applied for the third time on the concatenated vectors $\mathbf{b_s}$ and $\mathbf{b_g}$. In this combined model, a vector $\mathbf{c}$ controls shape and texture of the model. While matching an image with the model, an optimal vector of parameters is searched by minimizing the difference between the synthetic images and the given one. After matching, a best fitting model is generated, which actually gives the exact locations of all the facial features.

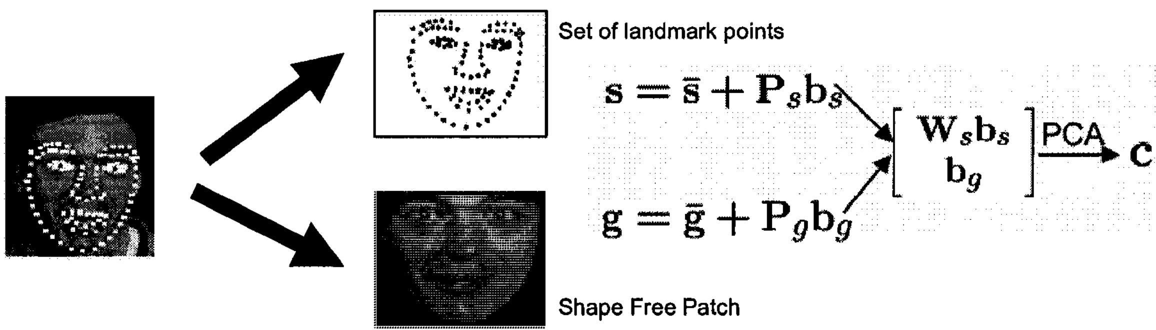

Figure 2.1: Active Appearance Model [3]

2.3 Face Recognition Approaches

Various face recognition techniques have been discussed in [1] and [2]. In the work of Zhao et al. face recognition methods are divided primarily into following three categories; table 2.1 shows the detail classification:

- Holistic Methods – these methods require the whole face region as the raw input.

- Feature Based or Structural Methods – for these approaches, the local features such as eyes, nose, mouth etc are extracted and their location and statistical information are fed to a structural classifier.
- Hybrid Methods – these methods use both the local features and the whole face region for recognition.

Table 2.1: Classification of Face Recognition Techniques [1]

Approach	**Representative Works**
Holistic Approaches	
Principal Component Analysis (PCA)	
Eigenface	Direct application of PCA
Fisherface	FLD on eigenspace
SVM	Two-class problem based on SVM
ICA	ICA based feature analysis
Other Representations	
LDA/FLA	FLD/LDA on raw images
PDBNN	Probabilistic decision based NN
Feature based methods	
Pure geometry methods	Earlier methods, recent methods
Dynamic Link Architecture	Graph Matching methods
Convolution Neural Network	SOM learning based CNN methods
Hybrid Methods	
Modular Eigenface	Eigenface & eigenmodules
Hybrid LFA	Local & global feature methods
Component based	Face region and components

2.3.1 Holistic Approaches

2.3.1.1 Face Recognition by PCA

PCA is a useful tool to obtain a lower dimensional representation of data. It is used not only to reduce statistical redundancy; but to reduce the system's sensitivity to noise (for example blurring, partial occlusion, change of background etc.) as well. Various face recognition algorithms have been proposed based on PCA. A brief overview of PCA is given here for ready reference [5, 6].

Let $X = \{X_n \in R^d \mid n = 1,2.....N\}$ be an ensemble of vectors, which is formed by converting each of the images to a vector by row concatenation, with d being the product of the width and height of an image. Let $E(x) = 1/N \sum_{n=1}^{N} X_n$ be the average vector in the ensemble. After subtracting the average from each element of X, we get a modified ensemble of vectors.

$$\overline{X} = \{\overline{X}_n, n = 1,2....N\} \text{ with } \overline{X}_n = \overline{X} - E(x) \tag{2.1}$$

The covariance matrix M for the ensemble X is defined by

$$M = \operatorname{cov}(\overline{X}) = E(\overline{X} \otimes \overline{X}) \tag{2.2}$$

where, M is a $d \times d$ matrix, with elements

$$M(i,j) = 1/N \sum_{n=1}^{N} [\overline{X}_n(i)\overline{X}_n(j)],\, 1 \le i,\ j \le d \tag{2.3}$$

It is well known from matrix theory that the matrix M is positively definite (or semidefinite) and has only real non-negative eigenvalues [5]. The eigenvectors of matrix M form an orthonormal basis for R_d. This is called Karhunen-Loeve (K-L) basis [7].

Sirovich and Kirby first used K-L transform to present human faces. Afterwards Turk and Pentland developed a PCA based face recognition system in 1991.

The work of Turk and Pentland [8] was motivated by that of Sirovich and Kirby. In their work, each of the training images was converted to a vector by row concatenation. The entire set of training images thus constructed the covariance matrix. This covariance matrix was solved for eigenvectors and eigenvalues. Then *n* best eigenvectors associated with *n* largest eigenvalues were selected; these eigenvectors were named 'eigenfaces'. This method claimed that each face can be represented as a weighted combination of selected eigenvectors or eigenfaces. The classification task was performed using nearest neighbour classifier. This method was found to be robust to illumination changes but performed weakly with scale variation.

Later, the work of Turk and Pentland was extended to a Bayesian approach by Moghaddam and Pentland in 1997 [9]; in their system the simple eigenspace based method was extended to use probabilistic measure of similarity. However, this scheme suffers from the problem of estimating probabilistic distributions in a high dimensional space from limited number of training examples per class [1]. In [10] Moghaddam argued Bayesian approach to be superior in terms of simplicity, computational efficiency and performance over PCA, ICA and nonlinear Kernel PCA (KPCA), where he examined these subspaces on the basis of experiments done on FERET database.

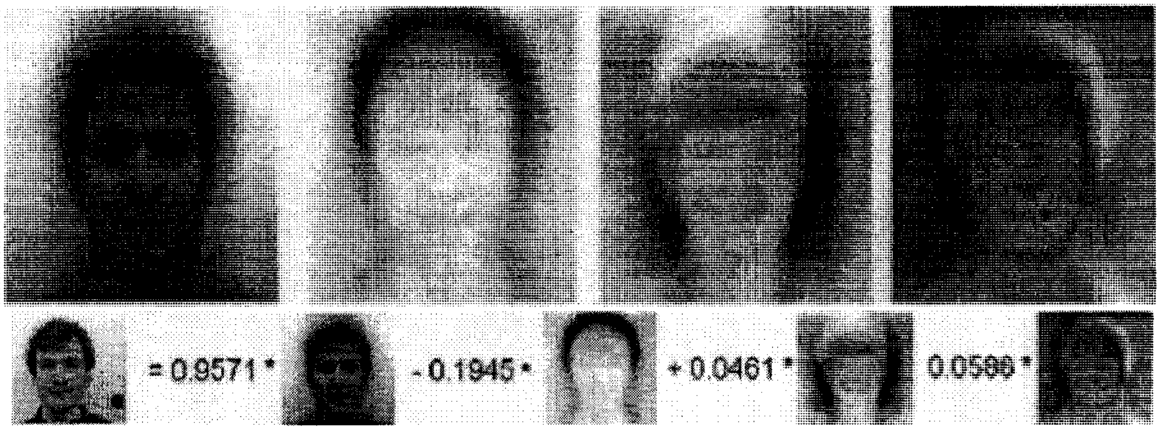

Figure 2.2: Eigenfaces [8]

PCA has now become one of the most popular tools for dimensionality reduction. Though this is widely used, it has two major drawbacks: high computational complexity and low discriminatory power. However, Chung et al. [11] suggested the use of Gabor Filter in combination with PCA in order to overcome the shortcomings of PCA. The authors of [11] argued that when raw images are subjected to PCA, the correlation of facial features is not well-reflected in eigenspace. So, they suggested using Gabor filter to extract facial features and then use PCA to classify the features optimally. It has been claimed that several problems like, deformation of face images due to in-plane in-depth rotation, illumination and contrast variation can be solved by extracting facial features using Gabor Filters.

2.3.1.2 Face Recognition by ICA

Independent Component Analysis is a generalization of Principal Component Analysis. ICA is argued to have the following advantages over PCA:

- ICA provides better probabilistic model of the data.

- ICA finds a not-necessarily orthogonal basis which can reconstruct the data better than PCA, even in the presence of noise.
- ICA is sensitive to higher order statistics of the data, not only covariance matrix.

Based on the above arguments, ICA has been used to extract features for face recognition by Bartlett et al. [12] in 1998 and performance improvement over PCA based method was reported.

2.3.1.3 Face Recognition by LDA

Linear Discriminant Analysis or Fisher Discriminant Analysis has also been successfully applied in face recognition by researchers [13]. LDA has been found to improve the classification accuracy of a system, when more than one image is available per class. LDA based recognition method aims at maximizing between-class scatter simultaneously minimizing within-class scatter. LDA training is performed via scatter matrix analysis [14]. For an N-class problem, the between and within class scatter matrix, S_b and S_w respectively, are calculated as follows:

$$S_b = \sum_{i=1}^{N} P_r(w_i)(n_i - n_0)(n_i - n_0)^T \tag{2.4}$$

$$S_w = \sum_{i=1}^{N} P_r(w_i) C_i \tag{2.5}$$

where $P_r(w_i)$ is prior class probability, n_i is conditional mean vector, n_0 is overall mean vector and C_i is average scatter of the sample vectors. Methods of combining PCA and LDA have been studied in [15, 16]. For this purpose first the face images are projected into eigenspace and then eigenspace projected vectors are again projected to LDA classification space before being subjected to classification.

2.3.1.4 Face Recognition by SVM

Support Vector Machine or SVM was employed as a classifier to recognize human faces by Phillips in 1998 [17]. Given a set of points, SVM finds the optimal 'hyperplane'. A hyperplane separates the data points into two classes and at the same time maximizes the distance from either class. This also indicates that SVM is a binary classifier whereas face recognition is a multi-class problem. The faces were projected in difference space where it can be considered as a two-class problem.

2.3.1.5 Neural Network Approaches

Neural Network based approaches have also been very popular in addressing the problem of face recognition. Neural networks have found their application in both holistic and feature-based approaches. Probability Decision Based Neural Network (PDBNN) method by Lin et al. [18] and Evolution Pursuit (EP) by Liu and Wechsler [19] are two successful examples of neural network approaches applied holistically. The PDBNN method is described here briefly.

PDBNN is an extension of DBNN proposed by Kung and Taur in 1995 [20]; PDBNN has three modules: face detector, eye localizer and face recognizer. Unlike other methods, this technique only considers the upper part of faces (eyebrows, eyes and nose, not mouth) in order to make the system robust to facial expression. The speciality of this method is its modularity i.e. for each class or person PDBNN dedicates one of its subnets for the representation of that particular person. Two sets of features are constructed from the segmented facial regions at reduced resolution and these two features are fed to two PDBNNs. The final identification is done by fusing the results of those two PDBNNs.

2.3.2 Feature Based Methods

Many feature based methods have been proposed till date. Some methods made use of geometry of local features, some used Hidden Markov Models (HMM) or Bayesian classifiers. Elastic Bunch Graph Matching (Okada et al. 1990 [21], Wiskott et al. 1997 [22]) is one of the most widely known algorithms of this kind.

2.3.2.1 Elastic Bunch Graph Matching

This technique is based on Dynamic Link Architecture or DLA (Buhmann et al. 1990 [23], Lades et al. 1993 [24]). DLA differs from traditional artificial neural network by the fact that syntactical relationships in neural network can be represented by DLA. Both Buhmann et al. and Lades et al. used Gabor wavelets to extract features (called 'jets'), which were locally estimated and were found robust to illumination changes, translation, distortion, scaling and rotation. 'Jets' [21] are small patches of grey values in an image around a pixel, which are based on wavelet transform and can be defined as a convolution of the image with a family of Gabor kernels [22]

$$\psi_j(\vec{x}) = \frac{k_j^2}{\sigma^2}\exp\left(-\frac{k_j^2 x^2}{2\sigma^2}\right)\left[\exp(i\vec{k}_j\vec{x}) - \exp(-\frac{\sigma^2}{2})\right] \tag{2.6}$$

A set of such jets (called 'bunch graph representation') was constructed from different facial images. EBGM based methods have been successfully applied in the areas of face detection, gender classification, pose estimation and object recognition.

2.3.2.2 Hidden Markov Models

HMM was applied for feature extraction and face identification by Samaria et al. in 1994 [25] and Nefian & Hayes in 1998 [26]. In this method the face images were intuitively divided into components such as forehead, eyes, nose, mouth and chin. These were considered as strips of pixels. Nefian et al. in 1998 reported better performance using KL projection instead of using raw pixel values. The recognition accuracy of HMM using ORL database [59] was found to be 87%. Later using 2D HMM [27] recognition performance was improved to 95%.

2.3.3 Hybrid Methods

2.3.3.1 Modular Eigenface

These methods use local features along with the entire face region. Modular Eigenface [28] by Pentland et al. 1994 is an example of hybrid methods where the global eigenfaces as well as the local eigenfeatures were used for identification. The concept of eigenfaces was extended to eigenfeatures, namely eigeneyes, eigenmouth etc. Recognition performance was computed as a function of eigenvectors, both for eigenfeatures alone and for combined representation. For lower-order spaces, eigenfeatures were found to be more effective compared to eigenfaces and the combination showed a marginal improvement only.

2.3.3.2 Flexible Appearance Model

Flexible appearance model [29] was used for automatic face recognition by Lanities et al. in 1995. Both shape and grey-level information had been used for recognition. The shape information was captured using Active Shape Model. To increase the robustness of the system against the presence of noise, scaling and partial occlusion, local grey-level model was also built on the shape model points. Finally, for an input image, three types of information, including shape parameters, shape-free image parameters and local profiles have been used for classification.

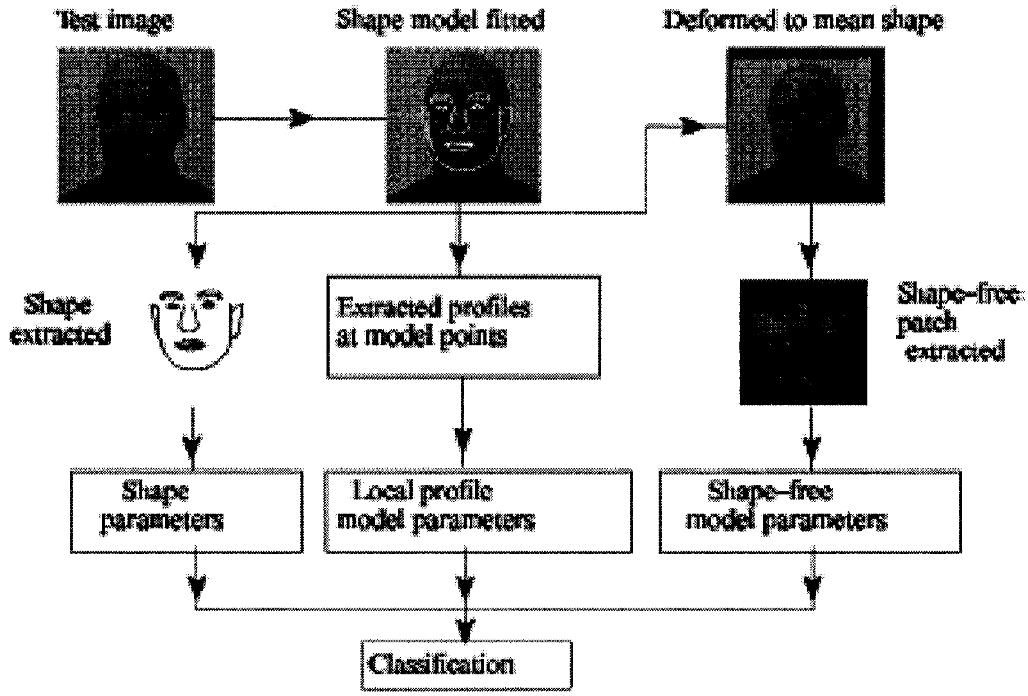

Figure 2.3: Flexible Appearance Model [1]

2.3.3.3 3D Morphable Models

Recently 3D Morphable models [30] by Blanz et al. 1999 and component-based recognition methods [31] by Heisele et al. 2001 have become quite popular. The key idea of component-based methods is to decompose the face into a number of facial components that are interconnected by a flexible geometric model. The system believes that by using components the pose change due to head movement can be accounted for. However, the method requires a large number of samples to work successfully.

3D Morphable Models have better ability to deal with the head pose and illumination variation problems. Unlike component based method it requires only three face images (frontal, semiprofile and profile) per person to compute the 3D face model. From this 3D face model synthetic images are generated. For face recognition nine facial components were used along with the entire face image, where as for face detection fourteen facial components were used. SVM was used as a classifier and recognition accuracy reported is quite high.

2.3.4 Other Approaches

Discrete Cosine Transform (DCT) and Fourier Transform have been applied to face recognition problem in [32] and [33] respectively. In [32] the authors tried to detect some critical areas of the facial images. The method was based on matching the image to a map of facial attributes associated with specific areas of face. DCT has also been used in combination with HMM in [34]. Spies and Ricketts in [33] used Fourier spectra for face analysis.

Most of the methods discussed above use frontal images. Liposcak et al. [35] have worked on profile images based on the original and morphological representation of derived profile shapes. They converted the grey-scale images to binary images. After normalization, they simulated the hair growth and haircut to derive two more profile silhouettes. Feature vectors extracted from these three profile images were used for recognition. 3D range images (obtained using laser scanner) have also been used as an input to face recognition system [2]. Infrared scanning of face images has been utilized in [36]. Some significant recent developments in face recognition field include 2D PCA [37], where Yang et al. used 2D image matrices for PCA instead of using 1D vectors, and face recognition using Laplacian-faces [38].

CHAPTER 3

WAVELET AND CURVELET TRANSFORM

3.1 Introduction

Multiresolution analysis tools such as Wavelet Transform have already been proved quite useful for analyzing the information content of images; and hence they have found their application in areas like image processing, pattern recognition and computer vision. This chapter discusses Wavelet Transform, the best known multiresolution tool and Curvelet Transform, a new multiresolution analysis tool used in this work. The underlying theory of both the transforms have been studied. The description of Wavelet Transform has been kept short as it is quite well-known and discussed in numerous books and research articles.

3.2 Wavelet Transform

Wavelet is a widely known multiscale transform, which is capable of providing the time and frequency information of a signal simultaneously. Fourier analysis of a signal enables us to represent a signal by the sum of a series of sines and cosines. But, Fourier expansion has only frequency resolution but no time resolution. Wavelet Transform, developed to overcome the limitations of Short Time Fourier Transform (STFT), solves this purpose and gives the time-scale representation of a signal [39].

In wavelet analysis a signal is processed at different scales or resolutions. The signal is cut up into different frequency components by a wavelet prototype function and each of the components is studied with a resolution matched to its scale. This wavelet prototype function $\psi(t)$ is called *'mother wavelet'*, as other wavelets are generated from this single basic wavelet by translation and scaling [40]. A wavelet function can be defined as a waveform of limited duration which has an average value of zero. As the original signal can be represented in a linear combination of such wavelet functions, further operations on the signal can easily be performed using just the corresponding wavelet coefficients.

It is worth mentioning that higher scale means more 'stretched' wavelets. More stretched the wavelet, longer the portion of signal it's being compared with, hence coarser the resolution. On the other hand lower scale corresponds to more 'compressed' wavelet, which in turn corresponds to rapidly changing details of the signal.

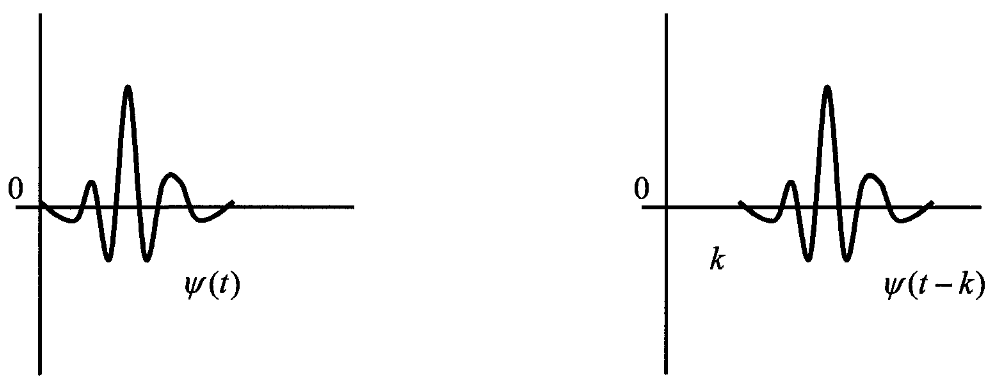

Figure 3.1: Shifting of Wavelet

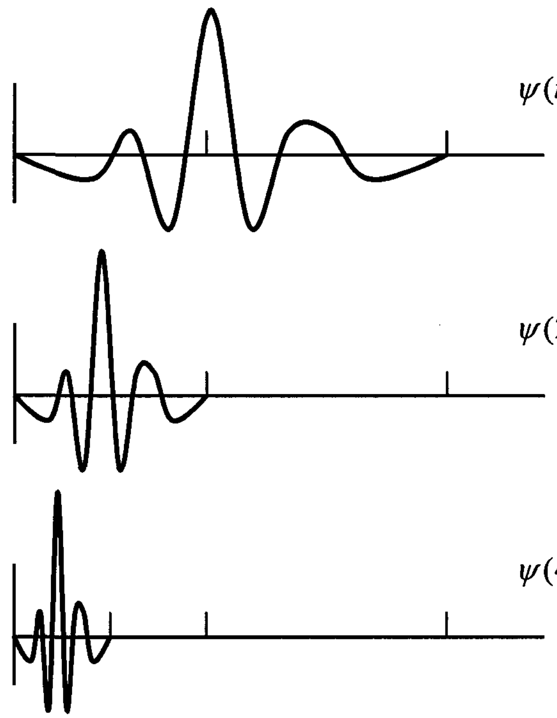

Figure 3.2: Scaling of Wavelet

3.2.1 Continuous Wavelet Transform

Continuous Wavelet Transform or CWT can be defined as the sum over entire time of the signal $f(t)$ multiplied by scaled, shifted versions of wavelet function $\psi(t)$ [39].

$$W(s,\tau) = \int_{-\infty}^{\infty} f(t)\psi^{*}_{s,\tau}(t)dt \tag{3.1}$$

Equation 3.1 shows how a signal $f(t)$ can be decomposed by a set of basis functions $\psi_{s,\tau}(t)$, called wavelets. Here, s and τ corresponds to scale and translation respectively. The wavelets $\psi_{s,\tau}(t)$ are generated from mother wavelet function $\psi(t)$ by scaling and translation as follows.

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right) \tag{3.2}$$

3.2.2 Discrete Wavelet Transform

CWT can operate at any possible scale; but computing wavelet coefficients at every possible scale is tedious and impractical. So some particular values of scale and position have to be chosen. Generally a didactic scale is chosen and using this, accurate results can be achieved. This is called Discrete Wavelet Transform. This is obtained by modifying the wavelet representation in (3.2) [39] as follows

$$\psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}}\psi\left(\frac{t-k\tau_0 s_0^j}{s_0^j}\right) \tag{3.3}$$

Here j and k are integers and $s_0 > 1$ is a fixed dilation step. τ_0 is the translation factor and depends upon dilation step.

3.2.3 2D Wavelet Decomposition

As this work deals with images i.e. 2D signals, 2D wavelet decomposition becomes an important topic of discussion. Wavelet decomposition can be regarded as the projection of a signal on the set of wavelet basis vectors. Two dimensional wavelet transform is derived from two 1D wavelet transform by taking tensor products [6]. It is implemented by applying 1D wavelet transform to the rows of the original (image data in this case) data and to the columns of the row transformed data. When wavelet transform is applied to an image it is decomposed into four subbands as shown in figure 3.3.

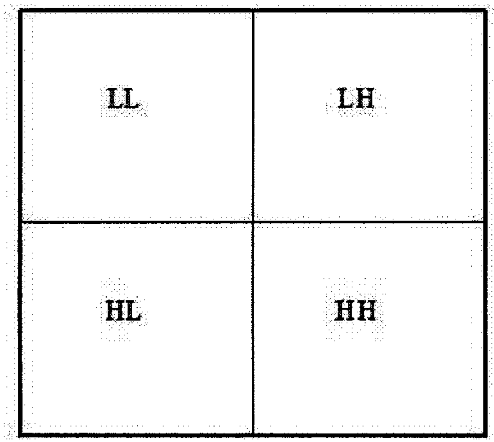

Figure 3.3: 2D 1-level Wavelet Decomposition

LL is a coarser approximation of the original image data. LH, HL and HH correspond to horizontal, vertical and high frequency changes of the data. Figure 3.3 shows 1-level decomposition. Further decomposition can be carried out on LL subband. Figure 3.4 shows a 3-level wavelet decomposition of the original 'lena' image (left). One of the advantages of wavelet decomposition is that, it provides local information in both frequency and space domain.

Figure 3.4: Left - Original Image; Right - 3-level Wavelet Decomposition [39]

3.3 Curvelet Transform

Over the past two decades, following the success of wavelets, other multiresolution tools like contourlets, ridgelets etc. were developed. "Curvelet Transform" is a recent addition to this list of multiscale transforms [41-43]. It was developed by Candes and Donoho in 1999. The development of curvelet transform was motivated by the need of image analysis [41]. This transform has improved directional capability, better ability to represent edges and other singularities along curves as compared to other traditional multiscale transforms, e.g. wavelet transform.

3.3.1 First Generation Curvelets

Curvelet construction is based on combining several ideas [41]:

- Ridgelets - a method of analysis suitable for objects with discontinuities along straight lines
- Multiscale Ridgelets - a pyramid of windowed ridgelets, renormalized and transported to a wide range of scales and locations
- Bandpass Filtering - separates an object out into a series of disjoint scales.

Curvelet Transform is based on above mentioned multiscale ridgelets. In combination with these, spatial bandpass filtering is used to isolate different scales. Curvelet coefficients are of two types: at coarse scale they are called 'wavelet scaling function coefficients' and at fine scale they are called 'multiscale ridgelet coefficients' of bandpass filtered object. Like ridgelets, curvelets can occur at any scale, location and orientation [42]. But curvelets have variable length in addition to variable width (variable anisotropy), where as ridgelets have only variable width and global length. At fine scale the relationship between width and length can be expressed as $width \approx length^2$ and anisotropy increases with decreasing scale obeying power law [42]. Figure 3.5 shows how curvelets look like at different scales, locations and orientations.

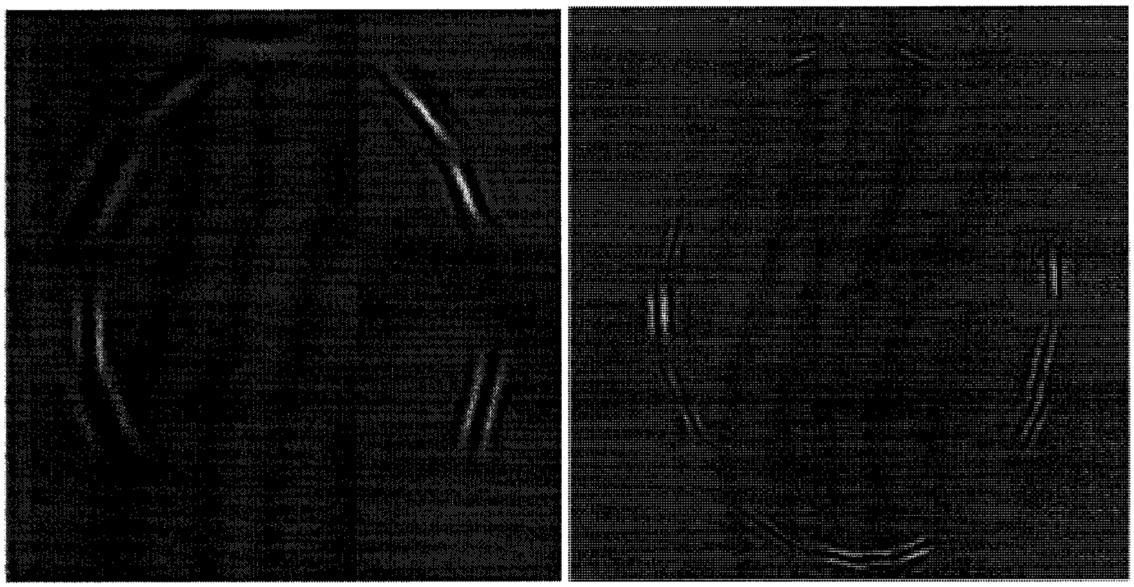

Figure 3.5: A Few Curvelets at different scale, position and angle [42]

3.3.2 Fast Discrete Curvelet Transform

In past few years curvelet construction has been redesigned in order to make it simpler to understand and use. This second generation curvelet transform [44] is not only simpler, but is faster and less redundant compared to its first generation versions discussed in previous section. In order to implement curvelet transform, first of all a 2D FFT of the image is taken. Then the 2D Fourier Frequency Plane is divided into 'parabolic' wedges. Finally an inverse FFT of each wedge is taken to find the curvelet coefficients at each scale j and angle ℓ. Figure 3.6 below shows the division of wedges of the Fourier Frequency plane in its left image; the right one represents curvelets in spatial Cartesian grid associated with a given scale and orientation [44]. From the right image, it can also be seen that curvelets are thin elliptical in shape; due to the parabolic relation between their width and length, they take the shape of elongated needles.

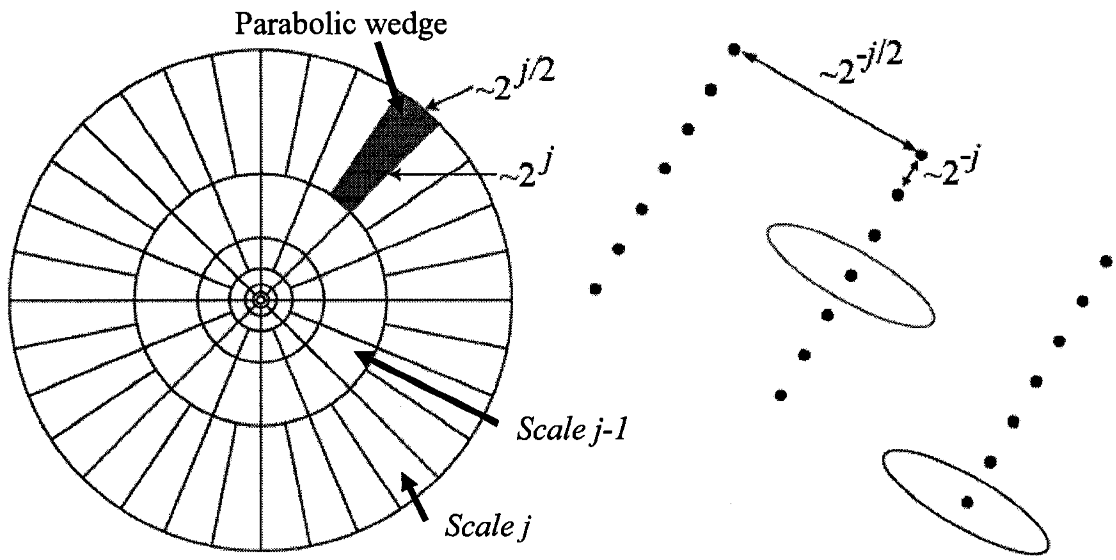

Figure 3.6: Curvelets in Fourier Frequency (left) and Spatial domain (right) [44]

There are two different digital implementations:

- Curvelets via USFFT (Unequally Spaced Fast Fourier Transform)
- Curvelets via Wrapping.

Though both the implementations use the same digital coronization but differ in the choice of spatial grid. Both of the FDCTs run in $O(n^2 \log n)$ flops for n by n Cartesian arrays [44]. Let $f[t_1, t_2], 0 \le t_1,\ t_2 < n$ be a Cartesian array and let $\hat{f}[n_1, n_2]$ denote its 2D discrete Fourier transform. Let $U_j(\omega)$ be a localizing window and $\widetilde{U}_j[n_1, n_2]$ is supported on some rectangle of length $L_{1,j}$ and width $L_{2,j}$ [44]

$$\mathrm{P}_j = \{(n_1, n_2) : n_{1,0} \le n_1 < n_{1,0} + L_{1,j}, n_{2,0} \le n_2 < n_{2,0} + L_{2,j}\} \quad (3.4)$$

Both the FDCT algorithms are described below.

Algorithm for FDCT via USFFT [44]:

1. Apply 2D FFT and obtain Fourier samples $\hat{f}[n_1, n_2]$, $-n/2 \le n_1,\ n_2 < n/2$.
2. For each scale/angle pair (j, ℓ), resample (or interpolate) $\hat{f}[n_1, n_2]$ to obtain sampled values $\hat{f}[n_1, n_2 - n_1 \tan\theta_\ell]$ for $(n_1, n_2) \in \mathrm{P}_j$.
3. Multiply the interpolated (or sheared) object $\hat{f}$ with the parabolic window $\widetilde{U}_j$, effectively localizing $\hat{f}$ near the parallelogram with orientation θ_ℓ, and obtain

$$\hat{f}_{j,\ell}[n_1, n_2] = \hat{f}[n_1, n_2 - n_1 \tan\theta_\ell]\widetilde{U}_j[n_1, n_2]. \quad (3.5)$$

4. Apply the inverse 2D FFT to each $\hat{f}_{j,\ell}$, hence collecting the discrete coefficients.

Algorithm for FDCT via Wrapping [44]:

1. Apply 2D FFT and obtain Fourier samples $\hat{f}[n_1, n_2]$, $-n/2 \le n_1,\ n_2 < n/2$.
2. For each scale j and angle ℓ, form the product $\widetilde{U}_{j,\ell}[n_1, n_2]\hat{f}[n_1, n_2]$.
3. Wrap this product around the origin and obtain

$$\hat{f}_{j,\ell}[n_1, n_2] = W(\widetilde{U}_{j,\ell}\hat{f})[n_1, n_2], \quad (3.6)$$

where the range n_1 and n_2 is now $0 \le n_1 < L_{1,j}$ and $0 \le n_2 < L_{2,j}$.

4. Apply the inverse 2D FFT to each $\hat{f}_{j,\ell}$, hence collecting the discrete coefficients.

3.3.3 Application of Curvelets

Curvelets, being a new concept, has not yet been very popular. So far, it has been successfully applied mostly in the fields of image processing, like image denoising [42], image compression [45], image fusion [46], contrast enhancement [47], image deconvolution [48], high quality image restoration [49], astronomical image representation [50] etc. Recently curvelets have also been employed to address a few problems of computer vision and patter recognition, like Optical Character Recognition [51], finger-vein pattern recognition [52] and palmprint recognition [53].

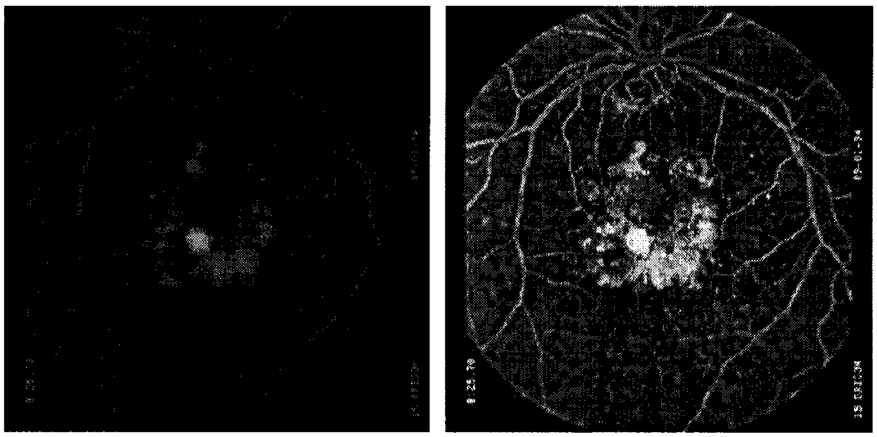

Figure 3.7: Contrast Enhancement by Curvelets [47]

Figure 3.8: Image Denoising by Curvelets Transform [42]

3.4 Wavelet vs. Curvelet

The sparsity of Fourier series is destroyed due to discontinuities (Gibbs Phenomenon); it requires a large number of terms to reconstruct a discontinuity in Fourier series within good accuracy. Later, wavelets are found to have the ability to solve the problems of Fourier series, as they are localized and multiscale. However, though wavelets do work efficiently in one-dimension, they fail to represent higher dimensional singularities effectively due to limited orientation selectivity. Wavelets and related classical multiresolution ideas exploit a limited dictionary made up of roughly isotropic elements occurring at all scales and locations [54]. These dictionaries do not exhibit highly anisotropic elements and there are only a fixed number of directional elements (The standard orthogonal wavelet transforms have wavelets with primarily vertical, primarily horizontal and primarily diagonal orientations) independent of scale. Images do not always exhibit isotropic scaling and thus these limitations of wavelets call for other kinds of multi-scale representation.

The most interesting fact about curvelets is that it has been developed specially to represent objects with *'curve-punctuated smoothness'* [54] i.e. objects which display smoothness except for discontinuity along a general curve; images with edges would be good example of this kind of objects. Wavelet transform has been profusely employed to address different problems of pattern recognition and computer vision because of their capability of detecting singularities. But, though wavelets are good at representing point singularities in both 1D and 2D signals, they fail to detect curved singularities efficiently. Figure 3.9 shows the edge representation capability of wavelet (left) and curvelet transform (right). For the square shape of wavelets at each scale, more wavelets are required for an edge representation than that compared to the number of required curvelets, which are of elongated needle shape. Roughly, to represent an edge of squared error $1/N$, $1/N$ wavelets but only $1/\sqrt{N}$ curvelets [55] are required. One more novelty of curvelet transform is that it is based on anisotropic scaling principal, where as wavelets rely on isotropic scaling.

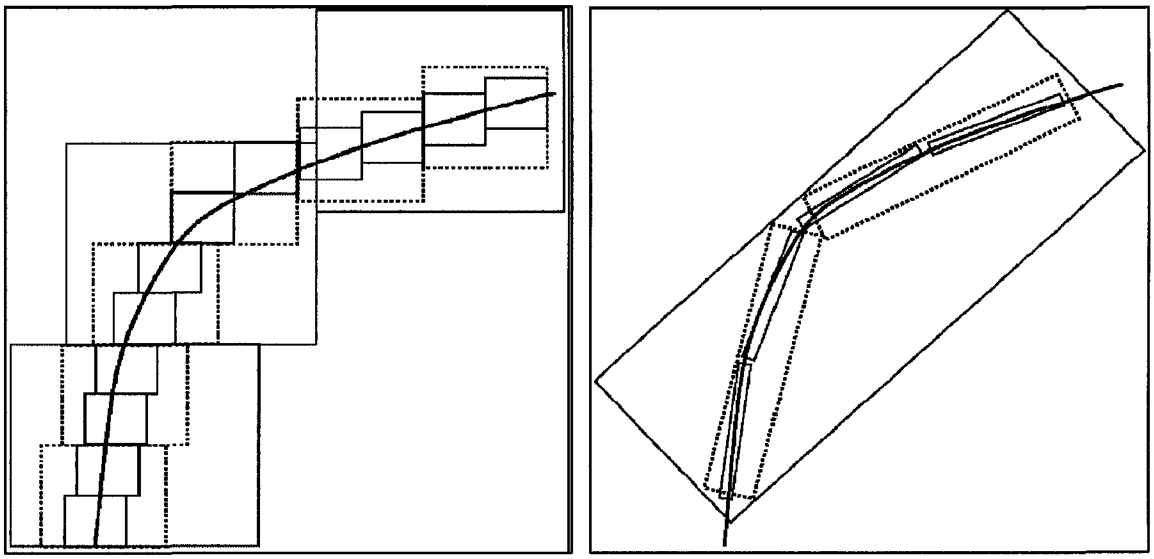

Figure 3.9: Edge Representation by Wavelet and Curvelet Transform [56]

Let us consider a function f, which has a discontinuity across a curve and is otherwise smooth; if it is approximated by the best m terms in Fourier expansion, the squared error is given by [56]

$$\left\| f - \tilde{f}_m^F \right\|^2 \propto m^{-1/2},\ m \to +\infty \tag{3.7}$$

For wavelets,

$$\left\| f - \tilde{f}_m^W \right\|^2 \propto m^{-1},\ m \to +\infty \tag{3.8}$$

For curvelets, we have

$$\left\| f - \tilde{f}_m^C \right\|^2 \propto \log(m)^3 m^{-2},\ m \to +\infty \tag{3.9}$$

To summarize, wavelet transform suffers from the following limitations:

- Edge representation - though wavelets perform better that FFT, it is not optimal.
- Limited number of directional elements independent of scale
- No highly anisotropic element

Curvelet transform is capable of solving the above problems. Curvelets thus can be considered as a higher dimensional generalization of wavelets and have the unique mathematical property that they can represent curved singularities effectively (hence the name) with very few coefficients and in a non-adaptive manner.

CHAPTER 4

THE CLASSIFIERS

4.1 Introduction

The task of assigning an object or event to one of several discrete categories (i.e. class) on the basis of prior knowledge is called 'classification'; and the task of a classifier is to partition feature space into class-labelled decision regions. Support Vector Machine (SVM) and *k*-Nearest Neighbour (*k*-NN) are the two classifiers that have been used in this work. This chapter contains brief overview of both the classifiers.

4.2 Support Vector Machine [57]

SVM models are a close cousin of classical neural networks. This is a supervised learning method, which determines the optimal hyperplane for linearly separable data and extends the patterns which are not linearly separable, to map into space by transformation of original data (by kernel function). The speciality of this classifier is that it always determines the hyperplane which maximizes the margin between two datasets. Using a kernel function, SVM is alternatively used as a training method for polynomial, radial basis function and multi-layer perceptron classifiers in which the weights of the network are found by solving a quadratic programming problem with linear constraints, rather than by solving a non-convex, unconstrained minimization problem as in standard neural network training.

The binary support vector classifier uses the discriminant function $f : X \subseteq \Re^n \rightarrow \Re$ of the following term

$$f(x) = \langle \alpha . k_s(x) \rangle + b \tag{4.1}$$

$k_s(x) = [k(x, s_1), k(x, s_2), ... k(x, s_d)]^T$ is the vector of evaluation of kernel functions centered at the support vectors $S = \{s_1, s_2, ... s_d\}, s_j \in \Re^n$, which are usually subset of the training data; α is the weight vector and b is a bias. The binary classification rule $q : X \rightarrow Y = \{1,2\}$ is defined as

$$q(x) = \begin{cases} 1 & for \quad f(x) \geq 0 \\ 2 & for \quad f(x) < 0 \end{cases} \tag{4.2}$$

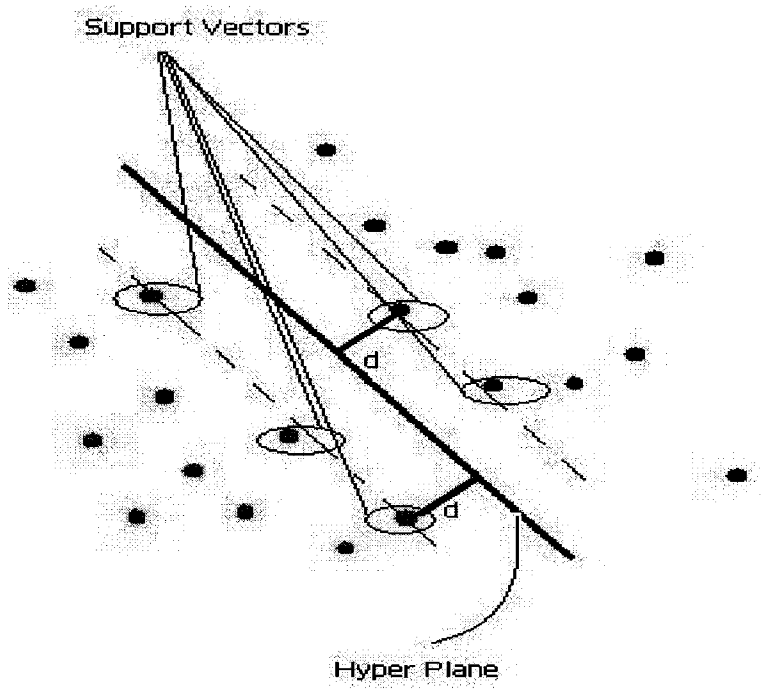

Figure 4.1: Support Vector Classification

4.2.1 Majority Voting with SVM

The majority voting strategy is a commonly used method to implement the multi-class SVM classifier. Let $q_j : X \subseteq \Re^n \to \{y_j^1, y_j^2\}$, $j = 1,2......g$ be a set of g binary SVM rules. The j^{th} rule $q_j(x)$ classifies the inputs x into class $y_j^1 \in Y$ or $y_j^2 \in Y$. Let $v(x)$ be a vector $[c \times 1]$ of votes for classes Y when the input x is to be classified. The vector $v(x) = [v_1(x), v_2(x)...... v_c(x)]^T$ is computed as

$$set \;\; v_y = 0 \,, \; \forall y \in Y$$

$$for \;\; j = 1,2....g \;\; do$$

$$v_y(x) = v_y(x) + 1 \,, \; where \;\; y = q_j(x);$$

$$end$$

The majority votes based multi-class classifier assigns the input x into such class $y \in Y$ having the majority of votes

$$y = \arg \max_{y^{'}=1,2......g} v_{y^{'}}(x) \tag{4.3}$$

4.2.2 One-Against-All and One-Against-One

The two most popular approaches are the One-Against-All (OAA) Method and the One-Against-One (OAO) Method. In this work an OAA SVM has been used; because it constructs g binary classifiers as against $g(g-1)/2$ classifiers required for OAO SVM while addressing a g class problem. The OAA decomposition transforms the multi-class problem into a series of c binary subtasks that can be trained by the binary SVM. Let the training set $T_{XY}^y = \{(x_1, y_1^{'}), ...(x_l, y_l^{'})\}$ contain the modified states defined as

$$y_i^{'} = \begin{cases} 1 & for \quad y = y_i \\ 2 & for \quad y \neq y_i \end{cases} \tag{4.4}$$

The discriminant functions

$$f_y(x) = \langle \alpha_y . k_x(x) \rangle + b_y, y \in Y \tag{4.5}$$

are trained by the binary SVM solver from the set T_{XY}^{y} , $y \in Y$.

4.3 *k*-Nearest Neighbor Classifier

Nearest Neighbor is an instance based learning method, where the key idea is to store all the training samples. Now given a query instance Y_q , the nearest neighbors of that query sample are found on the basis of distance; for k-nearest neighbor classifier, *k* such examples closest to Y_q are found. Then a majority voting scheme is used to determine the class of Y_q and it is assigned to the most common class among its *k* nearest neighbors.

4.3.1 Properties of *k*-NN

- This classifier does not build any model or try to simplify the dataset. It is more like a black box predicting the class of the query sample.
- Requires entire training dataset to be stored, leading to expensive computation if the data set is large. This effect can be offset, at the expense of some additional one-off computation, by constructing tree-based search structures to

allow nearest neighbors to be found efficiently without doing an exhaustive search of the data set [58].

- When $N \to \infty$, where N is the number of training samples, the error rate is never more than twice the minimum achievable error rate of an optimal classifier i.e. one that uses true classification distributions [58].
- This is robust to the noise present in training data; even removal of random points has little effect on its performance.
- Computational complexity is $O(kNd)$, where, N is the total number of training examples and $O(d)$ is the order of complexity to calculate distance to one training example.

4.3.2 Distance Measures

Usually, the distance between the query and any training example is computed in terms of Euclidean Distance. Other popular distance measures are Manhattan and L^n norm. Therefore the distance between $Y_q = (y_1, y_2 \ldots\ldots y_n)$ and $X = (x_1, x_2, \ldots\ldots x_n)$ is given by

Euclidean: $$D_{X,Y_q} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{4.6}$$

Manhattan: $$D_{X,Y_q} = \sqrt{\sum_{i=1}^{n} | x_i - y_i |} \tag{4.7}$$

L^n-norm: $$D_{X,Y_q} = \sqrt{\sum_{i=1}^{n} | x_i - y_i |^n} \tag{4.8}$$

4.3.3 How to Choose k

The most critical task here is to find the optimum value for *k*. For large training sets, large value of *k* seems to work better. But if *k* is too large the classifier becomes insensitive to noise. Small value of *k* might be necessary for small sets of training examples; In practice, k should be large enough so that error rate can be minimized but not too large, as that leads to over-smooth decision boundaries; k should also be kept small in order to include only the nearby samples, but again too small k value seems to lead to noisy boundaries. It is difficult to balance and find out an optimal value. k = 1 (nearest neighbour rule) is often used in practice. The following figure is an example taken from [58]; it is a plot of 200 data points from oil data set, showing values x_6 plotted against x_7. Figure 4.2 shows that value of *k* controls degree of smoothness of the decision boundary. Small *k* value produces may small regions for each class, but larger *k* results fewer and larger regions.

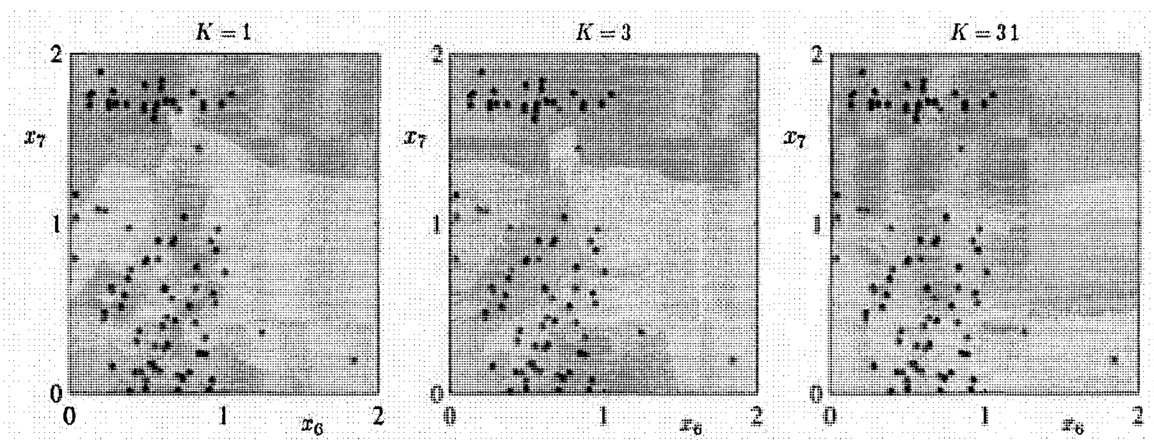

Figure 4.2: Effect of *k* Value [58]

CHAPTER 5

FACE RECOGNITION USING CURVELET FEATURES

5.1 Introduction

A thorough discussion on curvelet and wavelet transform is provided in chapter 3 according to which curvelets can be as good as (or, even superior to) wavelets for extracting features from facial images in section 3.4. In this chapter the results of different experiments carried out on three different datasets will be presented and analyzed along with comparative studies with wavelet-based techniques.

5.2 Datasets

Experiments have been carried out on three popular datasets: AT&T (ORL) database, Georgia Tech database and Essex Grimace database. A brief description of each of them is given below.

5.2.1 AT&T "The Database of Faces" [59]

This database contains 10 different images of each of 40 distinct subjects taken between April 1992 and April 1994 at Olivetti Research Laboratory, Cambridge, UK. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The images are in '.pgm' format and of dimension 92 x 112 (width x height), 8-bit grey levels.

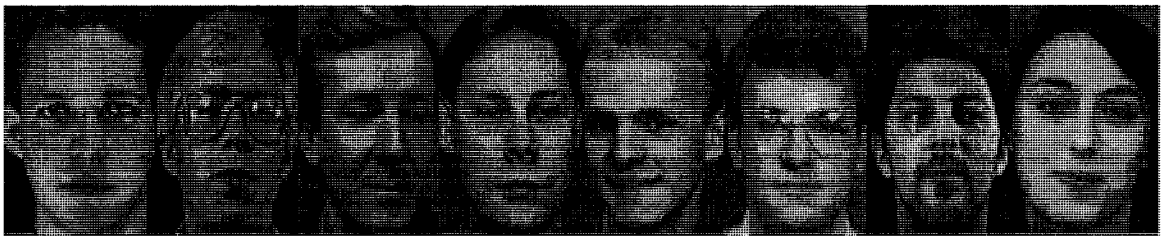

Figure 5.1: Sample images from AT&T database

5.2.2 Georgia Tech Database of Faces [60]

This database contains images of 50 people taken in two or three sessions between 06/01/99 and 11/15/99 at the Center for Signal and Image Processing at Georgia Institute of Technology. All people in the database are represented by 15 color JPEG images with cluttered background taken at resolution 640 x 480 pixels. The average size of the faces in these images is 150 x 150 pixels. The pictures show frontal and/or tilted faces with different facial expressions, lighting conditions and scale. The color images are converted to grey-scale images for experimental use. No other pre-processing like cropping or enhancing has been done.

Figure 5.2: Sample images from Georgia Tech database

5.2.3 Essex Grimace Database [61]

A sequence of 20 images each for 18 individuals consisting of male and female subjects was taken, using a fixed camera. During the sequence the subjects move their head and make grimaces which get more extreme towards the end of the sequence. There is about 0.5 seconds between successive frames in the sequence. Images are taken against a plain background, with very little variation in illumination. The images are in '.jpg' format and of size 180 x 200.

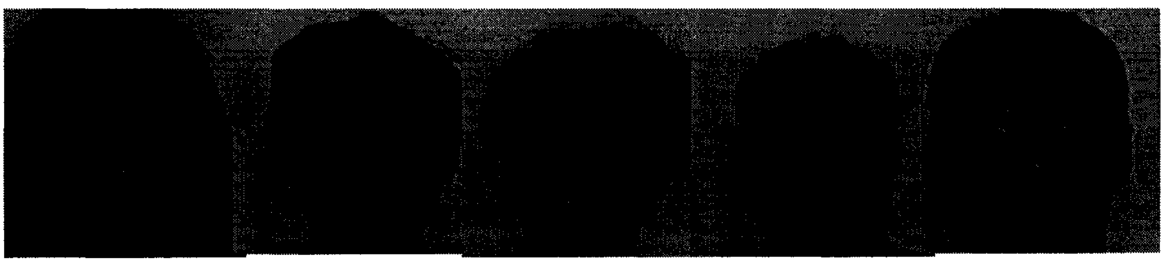

Figure 5.3: Sample images from Essex Grimace database

5.3 Curvelet Based Feature Extraction

Multiresolution analysis tools are applied in order to lessen the effect of variable facial appearance (expression variation, illumination variation, facial detail variation etc.)

on the classification systems. Also, working with high-dimensional images is computationally expensive. For these reasons, multiresolution analysis has gained immense popularity in the field of image processing and computer vision. The most popular multiresolution analysis tool is Wavelet Transform. It has enjoyed a wide-spread popularity in the field of computer vision. It has been profusely employed to address the problem of face recognition for feature extraction. Following the success of wavelets, a number of multiresolution analysis tools were developed in recent past; namely, contourlet, ridgelet and curvelet transform. In chapter 3 it has been argued theoretically, that curvelet transform, because of its better directional elements and highly anisotropic scaling property might supersede wavelets as the bases for face recognition. The following sections validate this empirically.

Our face recognition system is divided into two stages, training stage and classification stage. In the training stage, curvelet transform is applied to decompose the images. An interesting fact about curvelet transform is that, it has been developed especially to represent objects with *'curve-punctuated smoothness'* [54] i.e. objects which display smoothness except for discontinuity along a general curve; facial images with edges are good examples of this kind of objects. Say, the images are represented by 8 bits i.e. 256 gray levels. In such an image two very close regions can have differing pixel values. Such a gray scale image will then have a lot of "edges" – and consequently the curvelet transform will capture this edge information. These curvelet coefficients are fed to a classifier, which performs the identification task. According to [62], the recognition accuracy does not degrade if the size of the image is reduced. Following this as a preprocessing step, the length and width of the images have been reduced by two or four times, depending on the size of the original image. Training images are selected randomly and rest is used as test set. For cross-validation, the entire recognition process is repeated three times for each database. Figure 5.4 shows one approximate and detailed curvelet coefficients at eight different angles, for a sample face from AT&T database.

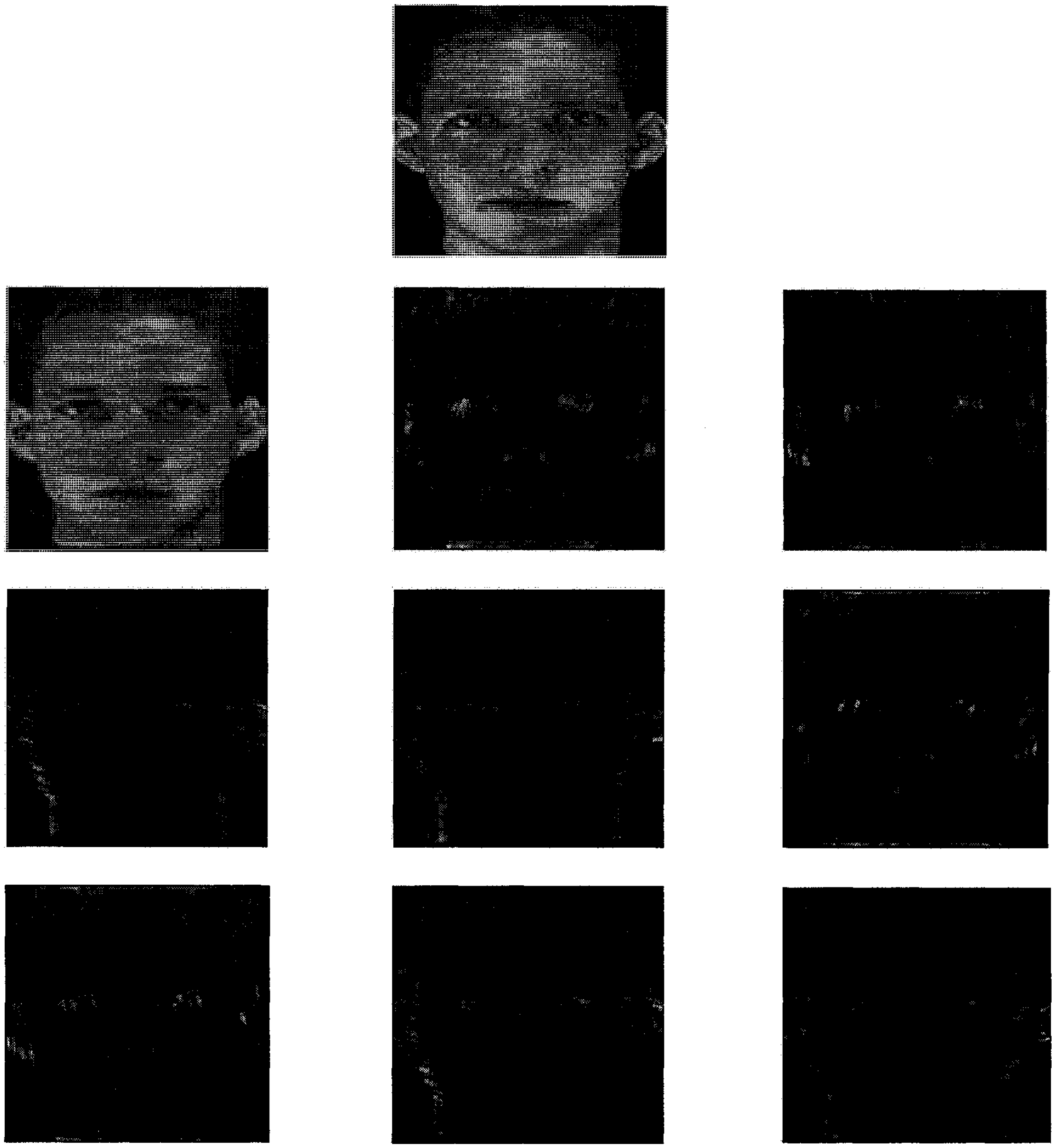

Figure 5.4: Curvelet transform of faces: first image is the original image, first image in the second row is the approximate coefficients and others are detailed coefficients at eight angles

All the experiments have been done with grayscale images only. The color images of Georgia Tech and Essex Grimace databases have been converted to grayscale images. Note that, the experiments have been carried out on three different databases to test the robustness of the algorithm against:

- Illumination variation - AT&T, Georgia Tech
- Expression variation - Essex Grimace
- Head movement - AT&T, Georgia Tech
- Scaling of subjects - Georgia Tech
- Cluttered background - Georgia Tech

5.3.1 Recognition using Simple Curvelet Features and *k*NN

This experiment is designed to work with simple curvelet features extracted from facial images. Due to lack of previous work in the area of curvelet transform for face recognition, this basic experiment is a good point to start from. In this experiment, the training images are simply subjected to curvelet transform and all the curvelet coefficients are fed to a k-NN classifier. L1 norm is used as the distance basis. Curvelet transform parameters are set to scale = 3 and angle = 8. These values for scale and angle are determined empirically, with the aim to balance recognition accuracy and computational efficiency of the system. Higher scale sometimes may result into better recognition rate (sometimes the improvement is only marginal), but that decreases the speed of the system as well.

There are three parameters for evaluating the performance of a face recognition system: recognition rate, rejection rate and false recognition or misclassification rate.

Recognition rate corresponds to the number of correct classification among all the images tested. The system rejects images in case of a tie or when no common class exists. When the images are assigned to the wrong class i.e. when Mr. A is wrongly identified as Mr. B, it is called false recognition. These parameters are usually expressed in terms of percentage. The results for the three datasets are given in the following page. Few additional results are listed in Appendix A.

5.3.1.1 Experiments on Essex Grimace Database

- Image size = 50 x 45
- Train: Test = 8:12

Table 5.1: Curvelet based results for Essex Grimace

Value of k	Recognition Rate (%)	Rejection Rate (%)	False Recognition Rate (%)
Set 1			
1	100	0.00	0.00
3	100	0.00	0.00
5	98.15	0.00	1.85
Set 2			
1	100	0.00	0.00
3	100	0.00	0.00
5	99.07	0.00	0.92
Set 3			
1	100	0.00	0.00
3	100	0.00	0.00
5	100	0.00	0.00

5.3.1.2 *Experiments on AT&T Database*

- Image size = 56 x 46
- Train: Test = 6:4

Table 5.2: Curvelet based results for AT&T

Value of k	Recognition Rate (%)	Rejection Rate (%)	False Recognition Rate (%)
Set 1			
1	93.13	0.00	6.87
3	87.50	8.12	4.37
5	78.75	9.37	11.87
Set 2			
1	93.87	0.00	6.12
3	86.87	8.75	4.37
5	76.25	13.12	10.62
Set 3			
1	94.37	0.00	5.62
3	87.50	6.25	6.25
5	80.00	9.37	10.62

5.3.1.3 Experiments on Georgia Tech Database

- Image size = 120 x 160
- Train: Test = 9:6

Table 5.3: Curvelet based results for Georgia Tech

Value of k	Recognition Rate (%)	Rejection Rate (%)	False Recognition Rate (%)
Set 1			
1	97.67	0.00	2.33
3	94.67	1.67	3.67
5	92.33	1.67	6.00
Set 2			
1	92.33	0.00	7.00
3	79.67	11.33	7.33
5	76.00	10.00	13.67
Set 3			
1	97.00	0.00	3.00
3	95.67	0.67	3.67
5	93.67	1.33	5.00

5.3.1.4 Comparison with Wavelets

In this section a quantitative comparison of the results achieved using curvelets and wavelets has been shown. 'Debauches' wavelet ('db1') has been employed to extract the wavelet coefficients. The same training and test sets have been used for wavelet based feature extraction keeping the image size same. The results in the following tables are achieved by averaging the results of three rounds of face recognition for both curvelet and wavelet based schemes.

Table 5.4: Comparison with wavelet based results on Essex Grimace

Value of k	Recognition Rate (%)	Rejection Rate (%)	False Recognition Rate (%)
		k = 1	
Curvelets	100	0.00	0.00
Wavelets	99.84	0.00	0.15
		k = 3	
Curvelets	100	0.00	0.00
Wavelets	99.69	0.00	0.31
		k = 5	
Curvelets	99.07	0.00	0.92
Wavelets	98.30	0.46	1.23

Table 5.5: Comparison with wavelet based results on AT&T

Value of k	Recognition Rate (%)	Rejection Rate (%)	False Recognition Rate (%)
		k = 1	
Curvelets	94.79	0.00	5.20
Wavelets	94.50	0.00	5.50
		k = 3	
Curvelets	87.29	7.70	4.99
Wavelets	85.83	10.00	4.16
		k = 5	
Curvelets	78.33	10.62	11.04
Wavelets	79.37	10.00	10.62

Table 5.6: Comparison with wavelet based results on Georgia Tech

Value of k	Recognition Rate (%)	Rejection Rate (%)	False Recognition Rate (%)
		k = 1	
Curvelets	95.88	0.00	4.11
Wavelets	98.00	0.00	2.00
		k = 3	
Curvelets	89.77	10.11	0.21
Wavelets	97.33	0.77	4.11
		k = 5	
Curvelets	87.67	4.11	8.22
Wavelets	95.11	0.77	4.11

5.3.1.5 Discussion

The experimental results of all the three databases are encouraging. Essex database shows extreme variation in facial expressions; high recognition accuracy achieved for this database (table 5.1) proves that curvelet features are robust to expression variation. Curvelet based features also seem to work well against illumination variation ad sidewise head movement as its performance is satisfactory for AT&T database. For these two datasets curvelets perform better than wavelets and thus support the theoretical argument made in chapter 3. However the accuracy achieved for Georgia Tech is lower compared to that achieved by wavelets (table 5.5 and 5.6). Cluttered background and scaling are the reason of this lower rate of recognition. Note that, no preprocessing operation like cropping, histogram equalization etc. have been performed on the images. Cropping the faces of Georgia Tech images and normalization, histogram equalization of images should improve the performance of the system.

5.3.2 Recognition using Bit Quantization, Curvelet Features and SVM

The previous experiment was based on simple application of curvelet transform for feature extraction to have a gross idea about its performance. This section introduces a novel idea of feature extraction using bit quantization. Features have been extracted by taking curvelet transform for each of the original image and its quantized 4 bit and 2 bit representations. The curvelet coefficients thus obtained act as the feature set for classification. These three sets of coefficients from three different versions of images are then used to train three Support Vector Machines. During testing, the results of these three SVMs (refer section 4.2) are fused using majority voting to determine the final classification result.

Black and white digital images are represented in 8 bits or 16 bits resulting into 256 or 65536 grey levels. All the image databases used here contain 8 bit images. If we quantize the grey levels to 128 or 64, nearby regions that had very little difference in pixel values and formed edges in the original 8 bit (256 grey levels) image will be merged and as a result only bolder edges will be represented. Now if these grey-level quantized images are curvelet transformed, the transformed domain coefficients will contain information of these bolder curves. The idea behind this scheme is that, even if a person's face is failed to be recognized by the fine curves present in the original 8 bit image, it may be recognized by bolder curves at a lower bit resolution. Images of the same persons from AT&T face database, quantized to 4 bits and 2 bits from the original 8 bit representation are shown in figure 5.5.

Figure 5.5: The images in the first column are the original 8 bit representations; the images in the second column are the 4 bit images while the last ones are 2 bit representations.

5.3.2.1 Experiments on Essex Grimace Database

- Image size = 45 x 50
- Train: Test = 12:8

Table 5.7: Curvelet based results for Essex Grimace

No of Bits	Recognition Rate of Individual SVM (%)	Recognition Rate after Majority Voting (%)	Rejection Rate (%)	False Recognition Rate (%)
		Set 1		
8	100			
4	100	100	0.00	0.00
2	100			
		Set 2		
8	95.83			
4	95.83	97.20	1.36	1.44
2	94.44			
		Set 3		
8	96.75			
4	95.67	97.20	2.10	0.90
2	95.67			

5.3.2.2 Experiments on AT&T Database

- Image size = 23 x 28
- Train: Test = 6:4

Table 5.8: Curvelet based results for AT&T

No of Bits	Recognition Rate of Individual SVM (%)	Recognition Rate after Majority Voting (%)	Rejection Rate (%)	False Recognition Rate (%)
Set 1				
8	96.93			
4	95.67	98.80	1.20	0.00
2	93.71			
Set 2				
8	98.82			
4	98.11	99.39	0.61	0.00
2	97.50			
Set 3				
8	96.93			
4	96.20	100	0.00	0.00
2	95.67			

5.3.2.3 Experiments on Georgia Tech Database

- Image size = 160 x 120
- Train: Test = 9:6

Table 5.9: Curvelet based results for Georgia Tech

No of Bits	Recognition Rate of Individual SVM (%)	Recognition Rate after Majority Voting (%)	Rejection Rate (%)	False Recognition Rate (%)
Set 1				
8	83.33			
4	83.33	88.70	7.00	4.30
2	79.72			
Set 2				
8	78.72			
4	78.00	81.30	10.59	8.11
2	74.00			
Set 3				
8	85.72			
4	85.33	89.70	5.70	4.60
2	84.33			

5.3.2.4 Comparison with Wavelets

To show how the curvelet based scheme fairs over wavelets, the same exercise as depicted in the previous sections has been carried out in the wavelet domain i.e.

wavelet transform has been used on bit quantized images. The results for these two schemes are compared in the following table for Georgia Tech database.

Table 5.10: Comparison with wavelet based results on Georgia Tech

No of Bits	Recognition Rate of Individual SVM (%)	Recognition Rate after Majority Voting (%)	Rejection Rate (%)	False Recognition Rate (%)
Set 1 – Curvelets				
8	83.33	88.70	7.00	4.30
4	83.33			
2	79.72			
Set 1 – Wavelets				
8	82.33	86.00	8.70	5.30
4	82.72			
2	78.00			
Set 2 – Curvelets				
8	78.72	81.30	10.59	8.11
4	78.00			
2	74.00			
Set 2 – Wavelets				
8	78.00	80.69	11.00	8.31
4	76.72			
2	73.33			
Set 3 – Curvelets				

8	85.72			
4	85.33	89.70	5.70	4.60
2	84.33			
Set 3 – Wavelets				
8	84.33			
4	84.00	87.33	6.27	6.40
2	83.73			

5.3.2.5 Discussion

From table 5.10, it is evident that curvelet based scheme provides better results against the wavelet based one, not only at the final level but for individual SVM performances as well. The technique appears to be robust to facial expression and illumination changes, as high recognition rate can be achieved for AT&T and Essex Grimace database. However, accuracy is lower for sidewise tilted images like those in Georgia Tech database. Recognition accuracy can be improved by cropping and making tilt corrections for the images of the Georgia Tech database. Other voting schemes can also be employed.

CHAPTER 6

FACE RECOGNITION USING CURVELET SUBSPACES

6.1 Introduction

The previous chapter discussed two new methods of feature extraction based on digital curvelet transform, where curvelet features were extracted from facial images or from their bit quantized versions. The results achieved were promising. In this chapter work will be done on image subspaces. Two dimension reduction tools, PCA and LDA will be applied on the curvelet features and the results will be discussed along with comparisons with existing techniques.

6.2 Datasets

The experiments for this work have been done on three well-known datasets: Essex Grimace database, Japanese Female Facial Expression (JAFFE) database and Faces94 database. Essex Grimace database has already been discussed in chapter 5, section 5.2.3. JAFFE and Faces94 are described in the following sections.

6.2.1 JAFFE Database [63]

This dataset contains 213 images of 7 facial expressions (6 basic facial expression and one neutral) posed by 10 Japanese female models. The images are taken against a homogeneous background and show extreme expression variation. The format of these images is '.tiff' and they are of size 256 x 256.

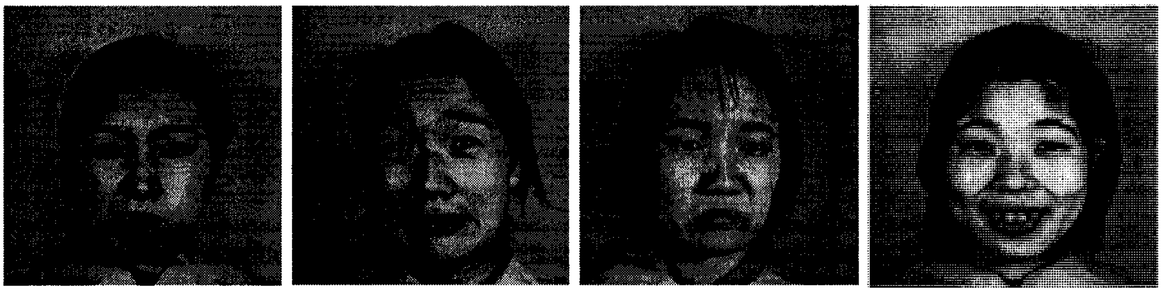

Figure 6.1: Sample images from JAFFE database

6.2.2 Faces94 Database [64]

This database contains images of 153 individuals (both male and female), 20 images per person. Subjects sit in front of a fixed camera and asked to speak. Images are contained in three different folders, male (113 individuals), female (20 individuals) and male staff (20 individuals). Faces of this database show considerable expression changes but no variation in lighting or head position. The background is plain green. The images are in .jpg format and of size 180 x 200. These color images are converted to grey-scale for the experiments.

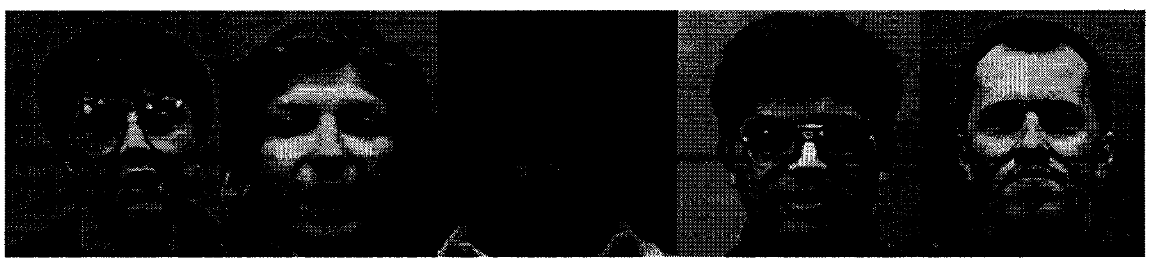

Figure 6.2: Sample images from Georgia Tech database

6.3 Recognition using Curvelet Subbands

In chapter 5 curvelet features extracted from facial images were directly used for classification. However, working with large number of features can be computationally expensive. So, reducing the dimensionality using linear analysis tools like PCA and LDA is necessary. The images are decomposed into its subbands using digital curvelet transform. Let's call them 'CURVELETFACES' (figure 6.3); these curveletfaces greatly reduce the dimensionality of the original image. Then PCA/LDA is applied on selected subbands. Thus an efficient and representative dataset can be produced. A simple Nearest Neighbour classifier is used to perform the classification task. The experiments have been designed specially to test the algorithm's robustness against expression variation and its ability to work on large datasets.

Figure 6.3 shows the curvelet subbands for a sample face taken from JAFFE dataset. Digital curvelet transform (scale = 2, angle = 8) was applied on the original image of size 256 x 256. This produced 1 approximate (171 x 171) and 8 detailed curvelet coefficients for 8 different angles. Throughout this work only the approximate (coarse) coefficients will be considered; hence that is our only concern. A scale value of 3 is chosen for this part of work in order to strike a balance between the speed and performance of the system. In our case, increment in scale parameter to 4 or 5 has shown only marginal improvement or even no improvement.

6.3.1 Recognition using Curvelet Subbands

PCA has been widely used in various problems of face recognition. Traditionally, each image is first converted to a vector by row (or column) concatenation.

Then, PCA is applied for dimensionality reduction. Though it provides effective approximation of data in lower dimension, the method suffers from high computational load and poor discriminatory power (refer section 2.3.1.1).

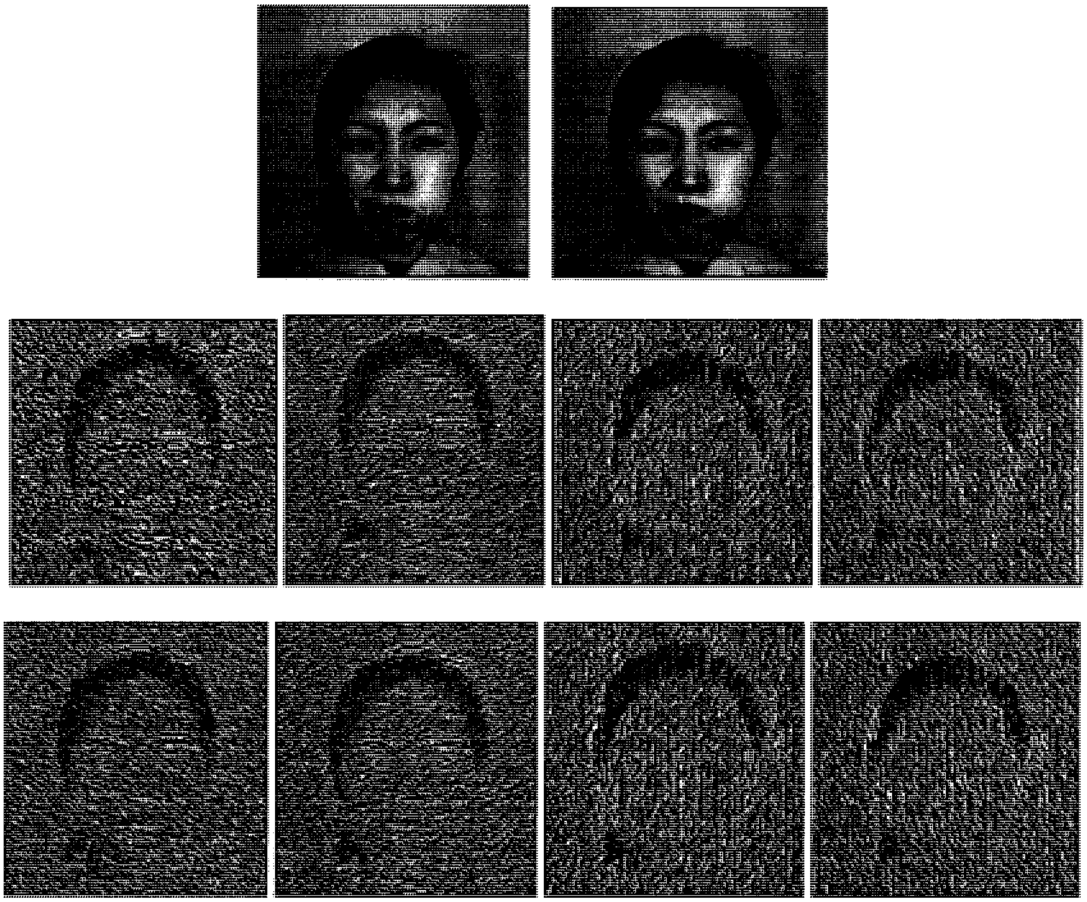

Figure 6.3: Curveletfaces: first image in the first row is the original image, second image in the first row is the approximate coefficients and others are detailed coefficients at eight angles

To solve these limitations researchers have applied PCA on wavelet subbands, where wavelet transform is used to decompose the face images in different frequency components; then a mid-range or low-range frequency subband is selected and subjected to PCA representation. This method has been proved to have better accuracy and better inter-class separability. Inspired by the success of subband PCA and superiority of curvelets over wavelets, PCA has been applied on curvelet subbands. The block diagram of figure 6.4 depicts the idea of our proposed method.

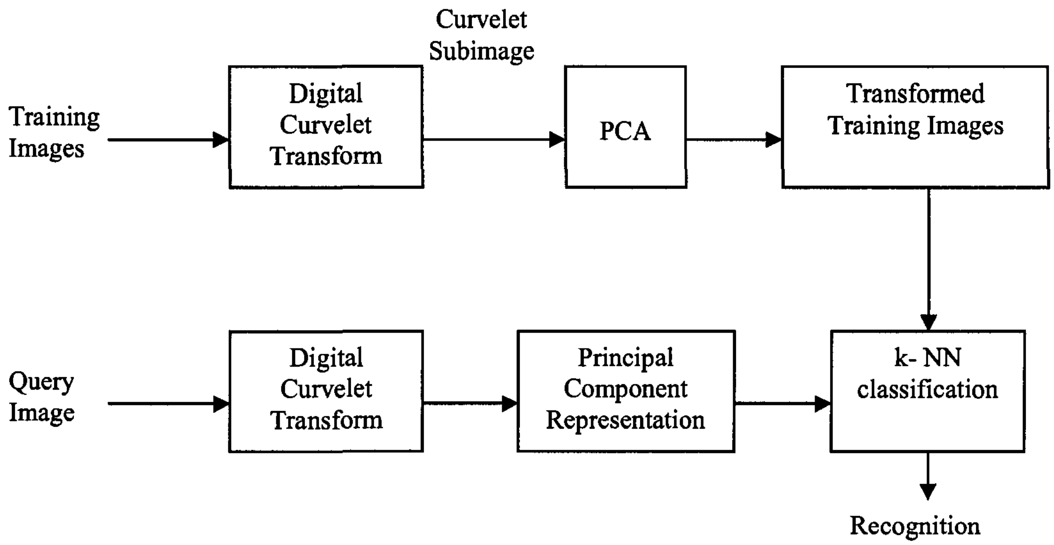

Figure 6.4: Curvelet based PCA face recognition scheme

PCA is applied on curvelet decomposed images for dimensionality reduction; a representational basis is formed. In classification stage the test images are subjected to the same operations and they are transformed to the same representational basis. A simple distance based classifier like k-NN classifier has been used to perform the classification task. Curvelet parameters are set to scale = 3 and angle = 8. At scale=3, the image is decomposed into 25 sub-bands, 1 approximate and 24 detailed sub-bands. As the approximate coefficients account for the maximum variance and contain most of the image energy, PCA has been applied on approximate coefficients only. This solves the problem of computational load. A 64 x 64 image when decomposed into curvelet subbands using scale = 3 and angle = 8, the resolution of the approximate sub-band will be 21 x 21 i.e. a vector of 121 features per image. For 50 x 45 image size the approximate curvelet subband will have a resolution of 17 x 15. The recognition accuracy is a function of number of eigenvectors. Different number of eigenvectors (principal components) has been selected to show the variation of recognition accuracy. Experiments have been carried out on Essex Grimace, JAFFE and Faces94 database with k = 1, 3 and 5. The recognition rates shown in the tables below are achieved by averaging the results of 5 different rounds of face recognition.

6.3.1.1 Results for Essex Grimace Database

- Image Size = 50 x 45
- Train : Test = 8 : 12

Table 6.1: Curvelets & PCA based results for Essex Grimace

No of Principal Components	Average Recognition Rate (%)		
	k = 1	*k = 3*	*k = 5*
5	99.54	99.38	97.22
10	99.84	99.69	98.14
15	99.84	99.69	98.30
20	99.84	99.84	98.45
25	99.84	99.84	98.76
30	99.84	99.84	98.76
40	99.84	99.84	98.92
50	100	99.84	98.92
70	100	99.84	98.92
90	100	99.84	98.92

6.3.1.2 Results for JAFFE Database

- Image Size = 64 x 64
- Train : Test = 9 : 13

Table 6.2: Curvelets & PCA based results for JAFFE

No of Principal Components	Average Recognition Rate (%)		
	k = 1	***k = 3***	***k = 5***
5	91.71	86.41	80.00
10	96.92	94.35	92.05
15	98.71	96.15	93.33
20	99.23	96.15	93.84
25	99.49	96.92	93.84
30	99.74	96.92	94.10
40	100	96.92	93.84
50	100	96.92	93.84
70	100	96.92	93.84
90	100	96.92	93.84

6.3.1.3 Results for Faces94 Database

- Image Size = 50 x 45
- Train : Test = 8 : 12

Table 6.3: Curvelets & PCA based results for Faces94

No of Principal Components	Average Recognition Rate (%)		
	k = 1	***k = 3***	***k = 5***
5	98.43	96.91	94.42
10	99.14	98.52	97.25
15	99.16	98.83	98.11
20	99.23	98.87	98.31
25	99.23	98.92	98.54
30	99.25	98.95	98.59
40	99.28	98.97	98.61
50	99.27	99.06	98.77
70	99.30	99.06	98.87
90	99.30	99.06	98.87

6.3.1.4 Comparison with Wavelet-based PCA and Traditional PCA

The tables in the previous three sections show how good curvelet features are in recognizing faces. However, these results do not have much credibility unless compared to some existing techniques. So, this proposed method has been compared with two popular existing techniques: wavelet-based PCA and traditional PCA. For wavelet-based PCA method, the images are subjected to multilevel wavelet (db1) subband decomposition and only the low frequency subimage (*ll* subband) is used for further operation. The same training and test sets are used for both wavelet-based and traditional PCA. The recognition accuracy is plotted against the number of principal components selected. The comparative results are shown graphically below in figure 6.5 and 6.6.

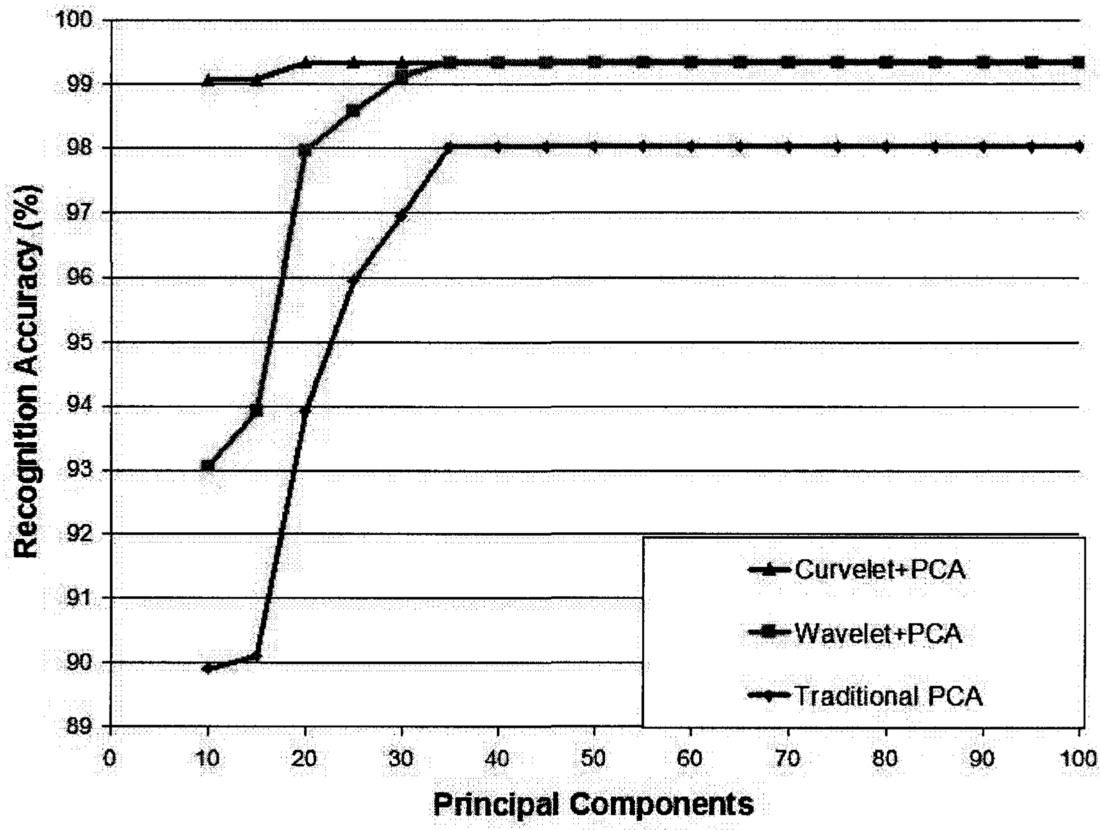

Figure 6.5: Comparison with wavelet based PCA and traditional PCA on JAFFE

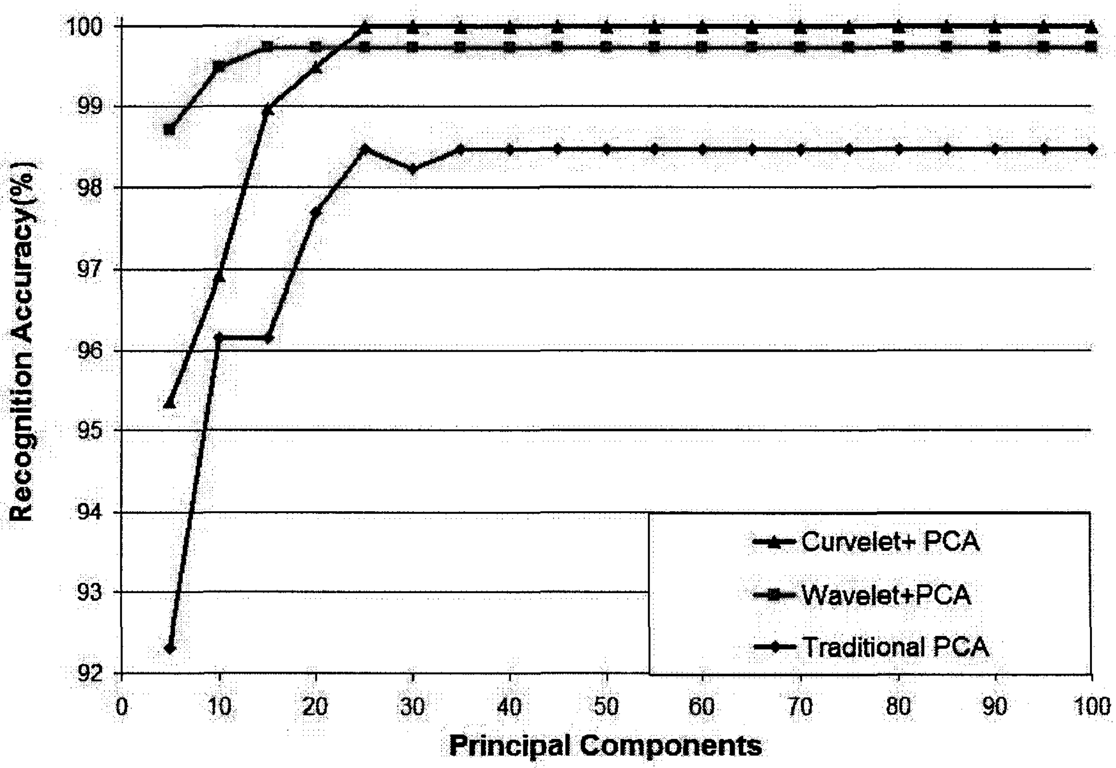

Figure 6.6: Comparison with wavelet based PCA and traditional PCA on Faces94

6.3.1.5 Discussion

From the previous graphical presentations in figure 6.5 and 6.6, it is evident that curvelet-based PCA performs considerably better than traditional PCA; but the performances of wavelet-based and curvelet-based methods are quite similar in terms of recognition accuracy. Our algorithm seems to be robust against facial expression changes and works effectively on large databases as it shows high recognition rate for the Faces94 database. These results suggest a new alternative approach to human face recognition, which is if not always better, as good as widely known wavelet-based PCA technique. Note that the purpose of this work is to explore the possibilities of curvelet transform in the field of face recognition. Theoretically curvelet transform has been claimed to be superior to wavelet transform in the literature; in this work, it has been verified empirically and compared with standard existing techniques, particularly for face recognition.

6.3.2 Recognition using Discriminant Curvelet Features

In section 6.3.1, PCA has been used as the dimensionality reduction tool. But researchers argue that though PCA provides an effective lower dimensional representation of data it suffers from poor discriminatory power. This problem can be solved by using Linear Discriminant Analysis or LDA (refer section 2.3.1.3), as it takes into account the between-class scattering. It aims at maximizing within-class similarity as well as between-class variance. Now LDA will be used for dimensionality reduction. A simple block diagram of the system architecture of this method is shown in figure 6.7.

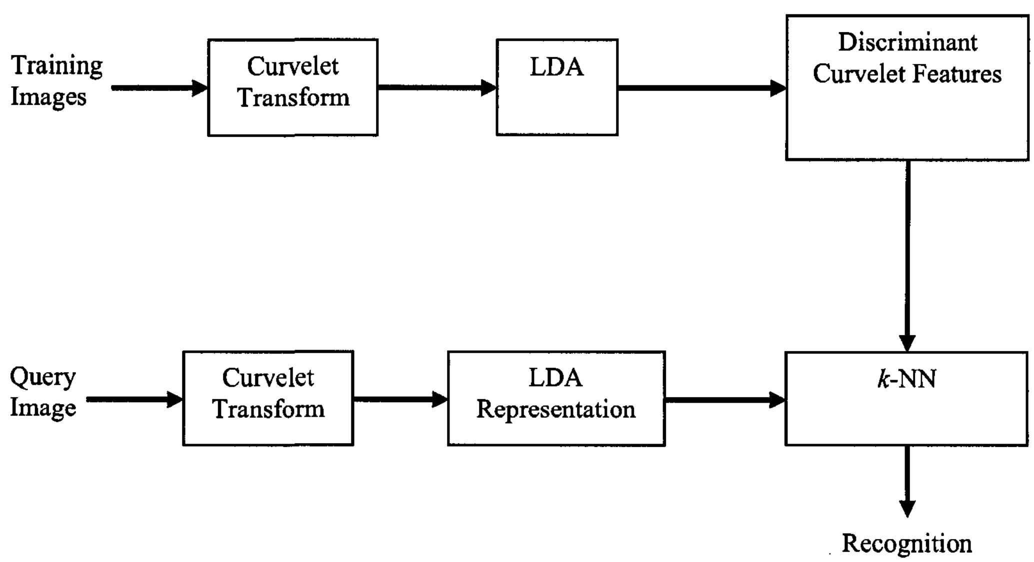

Figure 6.7: Recognition Scheme for Discriminant Curvelet Features

After decomposing the images using curvelet transform at scale = 3 and angle = 8, LDA has been applied on the resulting sub-images or curveletfaces. For Essex and Faces 94 subimage size is 17 x 15; however, for JAFFE images if curvelet transform at scale = 3 is used, the subband size will be 21 x 21 i.e. a feature vector of size 441; where, number of training images available for JAFFE database is around 100. For this situation S_w becomes singular. In order to solve this difficulty, curvelet transform at scale = 4 has been used for JAFFE database, which creates an approximate curveletface of dimension 11 x 11. To further reduce the dimensionality and enhance the between-class separability. A matrix $S = S_w^{-1} S_b$ is derived, where S_w is the within-class scatter matrix and S_b is the between-class scatter matrix. Then n numbers of eigenvectors corresponding to largest n eigenvalues are selected. During the experiments the number of eigenvectors (components) has been varied to study the variation in recognition accuracy.

A k-NN classifier is employed to perform the identification task like earlier experiments. The results for different databases are listed in following sections. The recognition rates listed in the tables are the results achieved by averaging the result of five different rounds of face recognition. Few additional results are presented in Appendix B.

6.3.2.1 Results for Essex Grimace Database

- Image Size = 50 x 45
- Train : Test = 8 : 12

Table 6.4: Discriminant curvelet based results for Essex Grimace

No of Principal Components	**Average Recognition Rate (%)**		
	k = 1	***k = 3***	***k = 5***
5	99.53	98.91	98.14
10	99.84	99.84	98.14
15	100	99.84	98.91
20	100	100	98.91
25	100	100	98.91
30	100	100	98.91
40	100	100	98.91
50	100	100	98.91
70	100	100	98.91
90	100	100	98.91

6.3.2.2 Results for JAFFE Database

- Image Size = 64 x 64
- Train : Test = 10 : 12

Table 6.5: Discriminant curvelet based results for JAFFE

No of Principal Components	**Average Recognition Rate (%)**		
	k = 1	***k = 3***	***k = 5***
5	95.89	92.81	88.97
10	96.68	94.35	91.79
15	97.17	94.35	92.30
20	98.89	96.69	92.30
25	98.89	96.69	92.30
30	98.89	96.69	92.30
40	100	96.69	93.07
50	100	98.69	93.07
70	100	98.69	93.07
90	100	98.69	93.07

6.3.2.3 Results for Faces94 Database

- Image Size = 50 x 45
- Train : Test = 8 : 12

Table 6.6: Discriminant curvelet based results for Faces94

No of Principal Components	Average Recognition Rate (%)		
	k = 1	***k = 3***	***k = 5***
5	98.40	97.45	96.38
10	99.16	98.62	98.13
15	99.23	99.06	98.30
20	99.23	99.06	98.35
25	99.23	99.06	98.90
30	99.23	99.06	99.06
40	99.23	99.17	99.06
50	99.79	99.17	99.06
70	99.79	99.17	99.06
90	99.79	99.17	99.06

6.3.2.4 Comparison with Discriminant Wavelet Method

In the previous section the results achieved using discriminant curvelet features has been presented. Now let us compare the results with wavelet-based results. The images are decomposed using multilevel wavelet transform ('db1'). Like previous PCA based experiment, only the low frequency *ll* subband has been selected for application of LDA. The comparative results are shown graphically below.

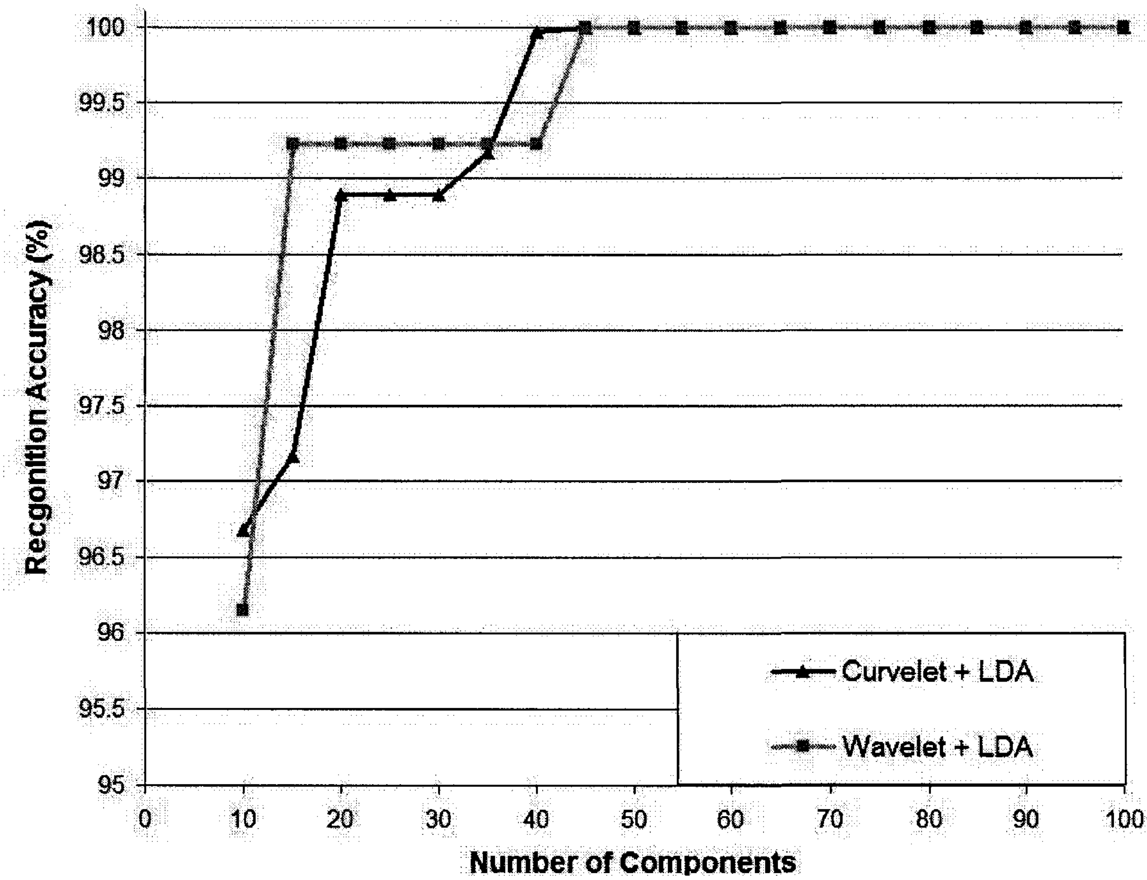

Figure 6.8: Comparison with discriminant wavelet based results on JAFFE

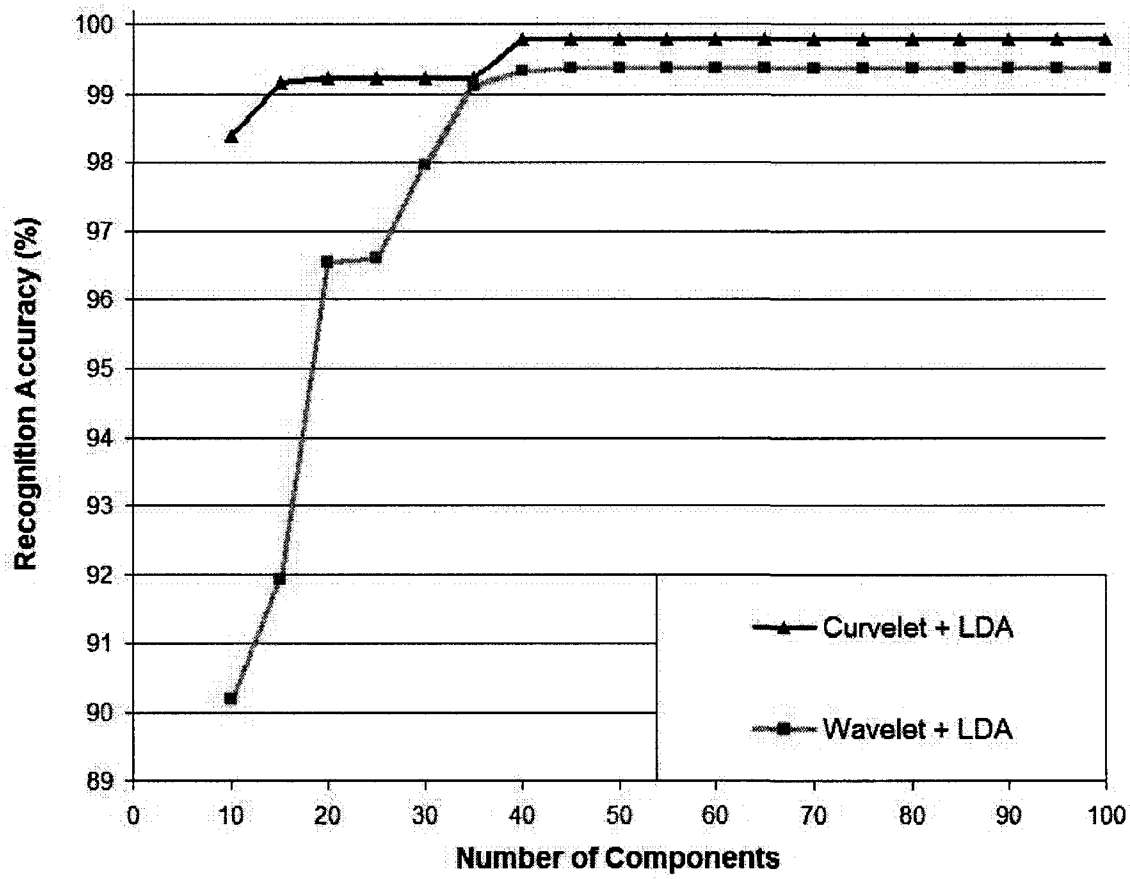

Figure 6.9: Comparison with discriminant wavelet based results fon Faces94

6.3.2.5 Discussion

*F*rom figure 6.8 and 6.9 it can be understood, that the performances of curvelet-based and wavelet-based LDA are comparable. This indicates that curvelet transform, when successfully combined with linear dimensionality analysis tools can serve as a good alternative to wavelet-based methods. For Faces94, curvelets show higher accuracy even for small number of components. While comparing curvelet-based PCA and curvelet-based LDA techniques, it has been found that the latter performs better.

CHAPTER 7

CURVELETS AND WAVELETS: A COMBINED APPROACH

7.1 Introduction

In chapter 5 and 6, we have discussed the potential of curvelet transform and compared its performance with that of its closest cousin, wavelet transform. So far we have shown curvelet transform as a good (if not always better) alternative to wavelet transform, for face recognition in particular. This chapter discusses a combined approach, which uses both curvelet and wavelet transform together in order to take advantage of both the transforms. This chapter also includes a comparative study of all the proposed methods with several existing methods.

7.2 Recognition using Curvelets and Wavelets

In chapter 6 the effects of using PCA and LDA on curveletfaces have been studied. The methods showed promising results, which is comparable to or sometimes even better than wavelet-based results. However, both the transforms have their own areas of expertise and weakness. Curvelets efficiently represent edge discontinuities but are challenged by small features, like eyes [65]. Again, wavelets though fail to capture the edge information faithfully, are very good at detecting point singularities. This method is based on the belief that these two transforms, if employed together will complement each other and thus will ensure higher rate of accuracy and increase the robustness of the system.

The face recognition system presented here not only combines curvelet and wavelet transform, but also uses a combined framework of PCA and LDA. The novelty of this method is the fusion of features extracted by two different mathematical transforms. The idea of combining PCA and LDA for dimensionality reduction has been investigated by many researchers as a popular framework for high dimensional data. LDA, though successfully applied for face recognition problems, encounters two major difficulties [66]. When the size of the training set is less than the dimensionality of images, the within-class scatter matrix becomes singular. Handling such large dimensional matrix can also be computationally expensive. To solve these difficulties LDA is often performed on a PCA transformed space. In the previous chapter, LDA was used on curvelet transformed images. There in order to avoid this problem of singular matrix, size of the original images were reduced beforehand; higher scale curvelet transform has been used as well. However, this new approach enables us to use the original images as well as solves the problem of high-dimensionality and singular scatter matrix.

In this method each image is subjected to both wavelet and curvelet decomposition in order to extract the intrinsic features of the facial images. 2D one-level wavelet transform decomposes an image into four subbands (refer chapter 3.2.3). Among which the lowest frequency sub-image (*ll* subband) is chosen and again decomposed using 2D one-level transform. After 3 levels of such decomposition the original 92 x 112 image produces a final low frequency *ll* subimage of dimension 12 x 13. Standard 'Haar' wavelet has been used to perform the decomposition. On the other hand, curvelet decomposition of the same image at scale 4 and angle 8 will produce 33 subbands (refer section 6.3); among these 33 subimages, only the approximate component which corresponds to the maximum variance is considered for future operation. Its dimension gets reduced to 19 x 15. Thus the dimensionality of the original image gets reduced by a large extent. To further reduce the dimensionality and reduce the statistical redundancy the wavelet and curvelet subimages are separately projected on eigenspace using PCA. Now on this PCA transformed space, LDA is performed; where there is no possibility of the within scatter matrix to be degenerate. Then equal number of LDA features extracted from both curvelet and wavelet decomposed images together form the final feature set for classification. The test images undergo a similar process and are converted to the same PCA + LDA representational basis. A k-Nearest Neighbour (k = 1) classifier is employed for identification. Figure 7.1 shows the architecture of this methodology.

7.2.1 Experimental Results

The evaluation of the proposed algorithm has been done on AT&T (ORL) database (refer section 5.2.1). Five images per person were selected randomly for prototypes and rest five per person were used to construct the test set. This means we have 200 training images and 200 test images. As mentioned before, after the original images are curvelet and wavelet transformed, they are projected to eigenspace from face space. Say, best 100 principal components are selected. This creates a feature vector of size 100 per image where the number of samples is 200. So, when LDA is applied, there is no possibility of the scatter matrix being degenerate. The results of this combined approach are compared with that of individual wavelet and curvelet schemes. Note that, equal number (say, 30 each) of LDA features have been selected from each of the curvelet and wavelet versions; the final number of features for the combined method is essentially doubled (i.e. 60). Recognition accuracy for three different rounds of face recognition is listed in table 7.1.

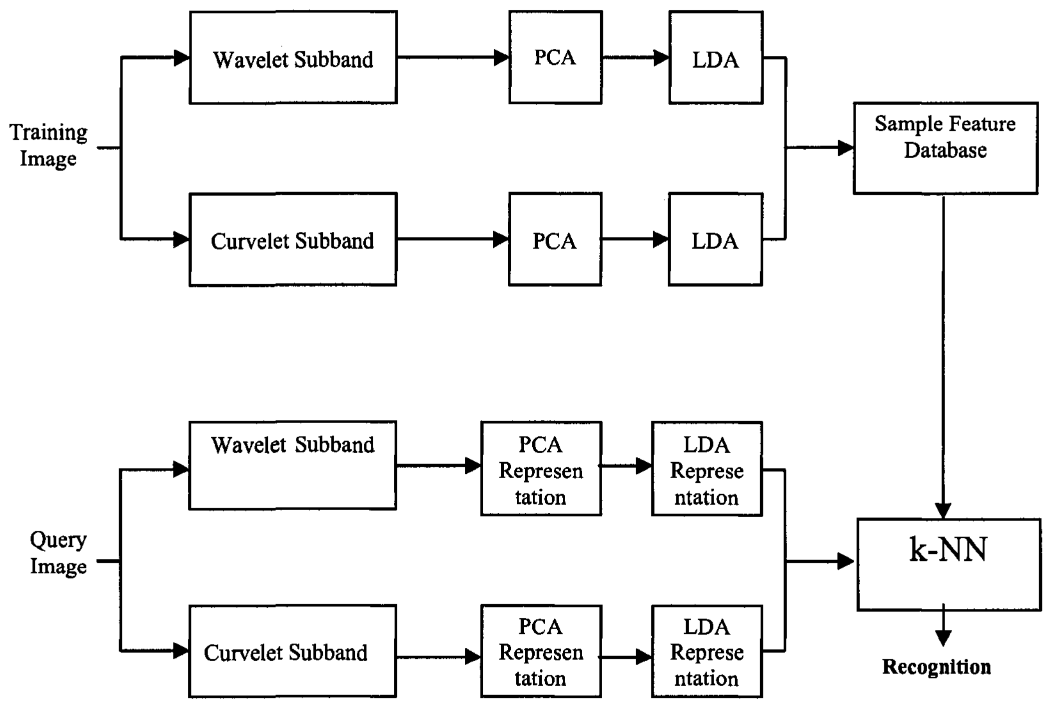

Figure 7.1: System Architecture

Table 7.1: Results for Combined Approach on AT&T

Components for Curvelet/Wavelet	Average Recognition Rate (%)		
	Combined Approach	Curvelet + PCA + LDA	Wavelet + PCA + LDA
Set 1			
10	95.00	94.00	93.00
25	96.50	96.00	95.50
35	96.50	95.00	95.00
Set 2			
10	93.50	92.50	93.00
25	95.00	94.50	94.50
35	95.50	93.50	93.50
Set 3			
10	96.00	95.50	93.50
25	97.50	96.00	95.50
35	97.50	96.50	95.50

7.2.2 Discussion

Table 7.1 shows that the fusion of two transforms provides better result than any of them alone. The results once again prove that curvelet based technique is comparable (even superior) to wavelet based technique. The drawback of this scheme, rather drawback of the PCA + LDA framework is that PCA criterion may not be compatible with LDA criterion [66]. Thus while performing PCA, some components may be ignored which had important discriminative information. To solve this problem at least partially, a large number of principal components have been chosen before performing LDA.

7.3 Comparative Study

A new idea of feature extraction for human face recognition using digital curvelet transform has been identified. Different schemes utilizing curvelet transform based feature extraction as well as a combined approach incorporating both curvelet and wavelet transform have been detailed so far. This section compares different curvelet based results with the results of Chien et al. [67] published in 2002, in order to evaluate the methodologies.

With the aim to make a fair comparison, same parameter values, mentioned by Chien et al. in [67] have been used. The dataset used is AT&T or ORL dataset (refer section 5.2.1). Among 10 images per individual, 5 have been randomly selected as prototype images and rest 5 construct the testing set, as mentioned in [67]. Following [55], selected number of components is 60; both for curvelet based PCA and curvelet based LDA methods. For the combined approach total number of final components is also 60 (30 each from curvelet and wavelet version). Image size has not been reduced this time. Curvelet transform is applied on the original images of ORL database of size 112 x 92. A 10 fold cross validation has been done to show statistical robustness of the methods. Simple *k*-Nearest Neighbour classifier with *k*=1 was used perform the recognition task. Final results have been compared in table 6.7.

Table 7.2: Comparative Study

Method	Recognition Accuracy (%)
Eigenface + k-NN [67]	92.00
Discriminant Eigenface + k-NN [67]	93.50
Waveletface + k-NN [67]	92.50
Curveletface + k-NN	**94.50**
Discriminant Waveletface + k-NN [67]	94.50
Discriminant Waveletface + MLP [67]	94.90
Curveletface + PCA + k-NN	**94.90**
Discriminant Waveletface + NFL [67]	95.10
Discriminant Curveletface + k-NN	**95.25**
Discriminant Waveletface + NFP [67]	95.80
Discriminant Waveletface + NFS [67]	96.10
Combined Approach	**96.75**

From the results listed in the above table, we summarize that curvelet based schemes outperforms wavelet based and eigenface methods. The experiments have been carried out on AT&T database, which shows significant illumination and expression variation. Even simple curvelet feature based method shows higher recognition accuracy compared to traditional eigenface, discriminant eigenface and waveletface method using kNN. Curvelet-based PCA technique has comparable results to that of discriminant waveletface using MLP. Discriminant curvelet features performs better than discriminant waveletface using NFL and finally, the combined approach supersedes most of the schemes shown in [67]. The combined approach achieves almost 5% higher accuracy than the standard eigenface technique and 4.25 % rise in accuracy compared to standard wavelet technique. Though the effect of different classifiers has not been studied, but the work of Chien et al. indicates that employment of classifiers like MLP, NFL, NFP and NFS should improve the performance of our proposed method highly. This suggests a scope of improvement and future work.

CHAPTER 8

CONCLUSION

8.1 Contribution

This thesis proposed a new approach to feature extraction from facial images using a new multiresolution analysis tool called digital curvelet transform. This mathematical tool though has been used in few image processing applications, has not yet been very popular among computer vision researchers as the concept of curvelet transform is quite new. Here we have identified digital curvelet transform as a feature extraction tool, which can be called a pioneer work.

As little work has been done with curvelet transform in computer vision field, it was required to start with the basic studies i.e. simple extraction of curvelet features. This showed promising results for curvelet based features. Then another idea of feature extraction from bit-quantized images was introduced. Inspired by the higher accuracy rate achieved by curvelet transform based methods, we went on to explore its possibility to be combined with linear analysis tools like PCA and LDA.

Two new feature spaces have been identified in this work. These new feature spaces are based on the PCA space and LDA space of the extracted curvelet features. In every step, the curvelet based results have been compared with wavelet-based methods for proper evaluation of the techniques. Then yet another novel idea of feature fusion was introduced. For this, both curvelet and wavelet transform based features have been used for subspace analysis, resulting into higher rate of recognition. Finally, the credibility and performance level of our proposed techniques was evaluated by comparing them with some high quality work published in prestigious journals [67]. As expected, proposed techniques have been found to be performing better.

The experiments have been designed to test the robustness of the algorithms mainly against expression and illumination variation, by selecting different datasets. We have been successful towards that too. This strongly indicates curvelet transform based feature extraction is promising for face recognition and can be successfully employed to address other aspects of this problem.

8.2 Possible Future Work

This work is an initiative to explore the different possibilities of curvelet transform for face recognition in particular; hence there is much scope of improvement and innovation. This work can be extended by investigating the effect of using other classifiers, cascading multiple classifiers and the performance using single training image. Other potential application areas could be emotion recognition, 3D face recognition, image retrieval etc. It can also be applied to address the problem of other biometric recognition like fingerprint or iris recognition. A preliminary work has been done on small-scale fingerprint recognition out of our interest in curvelet transform; a brief description can be found in Appendix C.

REFERENCES

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

[2] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey", *Proc. IEEE,* vol. 83, pp. 705-740, 1995.

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active Appearance Models", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681-685, 2001.

[4] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active Shape Models – Their Training and Application", *Computer Vision and Image Understanding*, vol. 61, pp. 18-23, 1995.

[5] A. K. Jain, *"Fundamentals of Digital Image Processing"*, pp. 163-175, Prentice Hall, New York, 1989.

[6] G. C. Feng, P. C. Yuen, and D. Q. Dai, "Human Face Recognition using PCA on Wavelet Subband", *Journal of Electronic Imaging*, vol. 9, no. 2, pp. 226-223, 2000.

[7] L. Sirovich and M. Kirby, "Low-dimensional Procedure for the Characterization of Human Faces", *Journal of Optical Society of America*, A 4, pp. 519-524, 1987.

[8] M. Turk and A. Pentland, "Face Recognition using Eigenfaces", *Proc. Computer Vision and Pattern Recognition*, pp. 586-591, 1991.

[9] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, pp. 696-710, 1997.

[10] B. Moghaddam, "Principal Manifolds and Probabilistic Subspaces for Visual Recognition" *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, pp. 780-788, 2002.

[11] K. C. Chung, S. C. Kee, and S. R. Kim, "Face Recognition using Principal Component Analysis of Gabor Filterresponses" *Proc. IEEE*, pp. 53-57, 1999.

[12] M. S. Bartlett, H. M. Lades, and T. Sejnowski, "Independent Component Representation for Face Recognition", *Proc. SPIE Symposium on Electronic Imaging: Science and Technology*, pp. 528-539, 1998.

[13] K. Etemad and R. Chellappa, "Discriminant Analysis for Recognition of Human Face Images", *Journal of Opt. Society Am.,* A 14, pp. 724-1733, 1997.

[14] K. Fukunaga, *"Statistical Pattern Recognition"*, Academic Press, New York, NY, 1990.

[15] W. Zhao, A. Krishnaswamy, R. Chellappa, D. L. Swets, and J. Weng, "Discriminant Analysis of Principal Components for Face Recognition", *Proc. International Conference on Automatic Face and Gesture Recognition*, pp. 336-341, 1998.

[16] W. Zhao, R. Chellappa, and P. J. Phillips, "Subspace Linear Discriminant Analysis for Face Recognition", *Tech. rep. CAR-TR-914,* Center for Automation Research, University of Maryland, College Park, MD, 1999.

[17] P. J. Phillips, "Support Vector Machines Applied to Face Recognition", *Adv. Neural Information Processing Systems*, vol. 11, pp. 803–809, 1998.

[18] S. H. Lin, S. Y. Kung, and L. J. Lin, "Face Recognition/Detection by Probabilistic Decision Based Neural Network", *IEEE Trans. Neural Network*, vol. 8, pp. 114–132, 1997.

[19] C. Liu and H. Andwechsler, "Evolutionary Pursuit and its Application to Face Recognition" *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, pp. 570–582, 2000.

[20] S. Y. Kung and J. S. Taur, "Decision-based Neural Networks with Signal/Image Classification Applications", *IEEE Trans. Neural Network*, vol. 6, pp. 170-181, 1995.

[21] K. Okada, J. Steffans, T. Maurer, H. Hong, E. Elagin, H. Neven, and C. V. D. Malsburg, "The Bochum/USC Face Recognition System and how it fared in the FERET Phase III Test", in *Face Recognition: From Theory to Applications*, H. Wechsler, P. J. Phillips, V. Bruce, F. F. Soulie, and T. S. Huang, Eds. Springer-Verlag, Berlin, Germany, pp. 186–205, 1998.

[22] L. Wiskott, J. M. Fellous, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, pp. 775–779, 1997.

[23] J. Buhmann, M. Lades and C. V. D. Malsburg, "Size and Distortion Invariant Object Recognition by Hierarchical Graph Matching", *Proc. International Joint Conference on Neural Networks*, pp. 411–416, 1990.

[24] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. V. D. Malsburg, R. Wurtz and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture", *IEEE Trans. Computers* ,vol. 42, pp. 300–311, 1993.

[25] F. Samaria and S. Young, "HMM Based Architecture for Face Identification", *Image Vision Comput.* , vol. 12, pp. 537–583, 1994.

[26] A. V. Nefian, and M. H. Hayes III, "Hidden Markov Models for Face Recognition", *Proc. International Conference on Acoustics, Speech and Signal Processing*, pp.2721–2724, 1998.

[27] S. Eickeler, S. Müller, and G. Rigoll, "High Quality Face Recognition in JPEG Compressed Images", *Proc. International Conference on Image Processing*, vol. 1, pp. 672-676, 1999.

[28] A. Pentland, B. Moghaddam and T. Starner, "View-based and Modular Eigenspaces for Face Recognition", *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1994.

[29] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic Face Identification System using Flexible Appearance Models", *Image Vision Comput.*, vol. 13, pp. 393-401, 1995.

[30] V. Blanz and T. Vetter, "A Morphable Model for the Synthesis of 3D Faces", *Proc. SIGGRAPH*, pp. 187–194, 1999.

[31] B. Heisele, T. Serre, M. Pontil, and T. Poggio, "Component-based Face Detection", *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[32] C. Podilchuk and X. Zhang, "Face Recognition Using DCT Based Feature Vectors", *Proc. IEEE ICASSP*, pp. 2144-2147, 1996.

[33] H. Spies and I. Ricketts, "Face Recognition in Fourier Space", *Proc. Vision Interface Conference*, pp. 38-44, 2000.

[34] S. Eickeler, S. Müller and G. Rigoll, "High Quality Face Recognition in JPEG Compressed Images", *Proc. International Conference on Image Processing*, pp. 672-676, 1999.

[35] Z. Liposcak and S. Loncaric, 1999, "Face Recognition from Profiles Using Morphological Operations", *Proc. International Conference on Information Technology*, 47-52.

[36] Y. Yoshitomi, T. Miyaura, S. Tomito and S. Kimura, "Face Identification Using Thermal Image Processing", *IEEE International Workshop on Robot and Human Communication*, pp. 374-379, 1997.

[37] J. Yang, D. Zhang, A. F. Frangi and J. Yang, "Two Dimensional PCA: a new approach to appearance-based face representation and recognition", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, pp.131-137, 2004.

[38] X. He, S. Yan, Y.Hu, "Face Recognition using Laplacianfaces", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, pp. 328-340, 2005.

[39] http://www.math.uiowa.edu/~jorgen/Haar.html

[40] MATLAB R2006A, Wavelet Toolbox documentation.

[41] E. J. Candes and D. L. Donoho, *"Curvelets- A Suprisingly Effective Nonadaptive Representation For Objects with Edges"*, Vanderbilt University Press, Nashville, TN, 2000.

[42] J. L. Starck, E. J. Candes and D. L. Donoho, "The Curvevet Transform for Image Denosing", *IEEE Trans. on Image Processing*, vol. 11, pp 670-684, 2002.

[43] E. J. Candes and D. L. Donoho, "Curvelets, Multiresolution Representation and Scaling Laws", *SPIE Wavelet Applications in Signal and Image Processing*, 2001.

[44] E. J. Candes, L. Demanet, D. L. Donoho, L. Ying. Fast Discrete Curvelet Transform, *SIAM Multiscale Model*, Simul. 2006.

[45] M. Manikandan, A. Saravanan, and K.B. Bagan, "Curvelet Transform Based Embedded Lossy Image Compression", *Proc. of ICSCN*, pp. 274-276, 2007.

[46] M Choi, R. Y Kim, and M. G. Kim, "The Curvelet Transform for Image Fusion", *Proc. ISPRS Congress*, Istanbul, 2004.

[47] J. L. Starck, F. Murtagh, E. J. Candes, and D. L. Donoho, "Gray and Color Image Contrast Enhancement by the Curvelet Transform", *IEEE Trans. Image Processing*, vol.12, pp.706-717, 2003.

[48] J. Starck, M. K. Nguyen, and F. Murtagh, "Deconvolution Based on the Curvelet Transform", *Proc. ICIP*, 2003.

[49] J. L. Starck, D. L. Donoho, and E. J. Candes, "Very High Quality Image Restoration by Combining Wavelets and Curvelets", *Proc. SPIE*, vol. 4478, 2003.

[50] J. L. Starck, D. L. Donoho, and E. J. Candes, "Astronomical image Representation by Curvelet Transform", *Astronomy & Astrophysics*, vol. 398, pp. 785-800, 2002.

[51] A. Majumdar, "Bangla Basic Character Recognition using Digital Curvelet Transform", *Journal of Pattern Recognition Research*, vol. 2, pp 17-26, 2007.

[52] Z. Zhang, S. Ma, and X. Han, "Multiscale Feature Extraction of Finger-Vein Patterns Based on Curvelets and Local Interconnection Structure Neural Network", *Proc. ICPR*, vol. 4, pp. 145-148, 2006.

[53] K Dong, G Feng, and D Hu, "Digital Curvelet Transform for Palmprint Recognition", *Lecture notes in Computer Science, Springer*, vol. 3338, pp. 639-645, 2005.

[54] E. J. Candes, "What is Curvelet", *Notices of American Mathematical Society*, vol. 50, pp. 1402-1403, 2003.

[55] D. L. Donoho and M. R. Duncan, "Digital Curvelet Transform: Strategy, Implementation and Experiments", *Technical Report*, Department of Statistics, Stanford University, 1999.

[56] J. L. Starck, "Image Processing by the Curevelet Transform", powerpoint presentation.

[57] C. Cortes and V. Vapnik, "Support-Vector Networks". *Machine Learning*, vol. 20, pp. 273–297, 2005.

[58] C. M. Bishop, *"Pattern Recognition and Machine Learning"*, Springer, 2006.

[59] http://www.cl.cam.ac.uk/Research/DTG/attarchive:pub/data/att_faces.zip

[60] http://www.anefian.com/gt_db.zip

[61] http://cswww.essex.ac.uk/mv/allfaces/grimace.zip

[62] J. Ruiz-Pinales, J. J. Acosta-Reyes, A. Salazar-Garibay, R. Jaime-Rivas, "Shift Invariant Support Vector Machines Face Recognition System", *Transactions on Engineering Computing and Technology*, vol. 16, pp. 167-171, 2006.

[63] http://kasrl.org/jaffe.html

[64] http://cswww.essex.ac.uk/mv/allfaces/faces94.html

[65] J. L. Starck, M. K. Nguyen, F. Murtagh, "Wavelets and curvelets for image deconvolution: a combined approach", *Signal Processing*, Vol. 83, Issue 10, pp. 2279-2283, 2003.

[66] H. Yu, J. Yang, "A direct LDA algorithm for high-dimensional data-with application to face recognition", *Pattern Recognition,* Vol. 34, Issue 10, pp. 2067-2070, 2001.

[67] J. Chien, C. Wu, "Discriminant Waveletfaces and Nearest Feature Classifiers for Face Recognition", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol 24, No. 12, pp. 1644-1649, 2002.

[68] Lecture Notes of 88-590(33), Machine Learning and Computer Vision, 2007

[69] http://www.curvelet.org

[70] http://www.eso.org/sci

APPENDIX A

Additional Results for Simple Curvelet Feature Based Recognition

Experiments on JAFFE Database

- Image size = 64 x 64
- Train: Test = 9:13

Table I: Curvelet Based Results for JAFFE

Value of k	Recognition Rate (%)	Rejection Rate (%)	False Recognition Rate (%)
Set 1			
1	98.46	0.00	1.53
3	97.69	0.77	1.53
5	96.15	1.53	2.30
Set 2			
1	100	0.00	0
3	97.69	0.76	1.53
5	95.38	1.53	3.07
Set 3			
1	99.23	0.00	0.77
3	93.84	0.77	5.38
5	95.38	0.77	3.84

Table II: Comparison with Wavelets on JAFFE

Value of k	Recognition Rate (%)	Rejection Rate (%)	False Recognition Rate (%)
k = 1			
Curvelets	99.23	0.00	0.00
Wavelets	98.46	0.00	1.53
k =3			
Curvelets	96.41	0.77	2.82
Wavelets	96.15	1.28	2.56
k = 5			
Curvelets	95.64	1.28	3.07
Wavelets	93.84	3.07	3.07

Experiments on Faces94 Database

Image size = 50 x 45

Train: Test = 8:12

Table III: Curvelet Based Results for Faces94

Value of k	Recognition Rate (%)	Rejection Rate (%)	False Recognition Rate (%)
Set 1			
1	99.23	0.00	0.76
3	98.84	0.22	0.93
5	98.57	0.32	1.09
Set 2			
1	99.12	0.00	0.8772
3	97.80	0.71	1.4803
5	97.58	0.43	1.9737
Set 3			
1	99.17	0.00	0.8224
3	98.95	0.21	0.8224
5	98.57	0.27	1.1513

Table IV: Comparison with wavelets on Faces94

Value of k	Recognition Rate (%)	Rejection Rate (%)	False Recognition Rate (%)
	k = 1		
Curvelets	99.17	0.00	0.84
Wavelets	98.77	0.00	1.23
	k =3		
Curvelets	98.53	0.38	1.09
Wavelets	97.14	0.67	2.18
	k = 5		
Curvelets	98.24	0.34	1.41
Wavelets	95.28	1.11	3.20

APPENDIX B

Additional Results for Discriminant Curvelet Feature Based Recognition

Experiments on AT&T Database

- Image Size = 56 x 46
- Train: Test = 5:5

Table V: Results for AT&T

No of Components	Average Recognition Rate (%)		
	k = 1	*k = 3*	*k = 5*
5	84.78	78.12	69.79
10	94.58	88.96	80.83
15	95.42	91.03	85.00
20	95.62	92.08	87.08
25	95.62	92.46	87.49
30	95.62	91.67	88.54
40	95.62	91.87	88.54
50	95.62	92.08	88.50
60	95.62	91.41	88.50
70	95.62	90.41	85.41
90	95.62	90.21	85.41

APPENDIX C

Idea of Fingerprint Recognition using Curvelet Transform

In order to find out the performance of curvelet transform for other biometric recognition, we have also done a small investigation on fingerprint recognition. A brief overview of curvelet based features extraction for fingerprint recognition is presented below. A few results are also shown in support.

Our method of curvelet based feature extraction partly follows the steps mentioned in the work of Tico et al[1]. The feature extraction steps used are as follows:

- Locate reference point or core point in the fingerprint image.
- Crop an N x M subimage from the fingerprint pattern with the core point at the center. This can be called central subimage[1].
- Divide the central subimage into a number of non-overlapping blocks of size W x W.
- Take curvelet transform for each of the non-overlapping blocks at Scale = S and Angle = A.
- Take standard deviation of each of the curvelet coefficient set for each scale and angle.
- The standard deviation of curvelet coefficients thus obtained for each block construct the global feature vector for a fingerprint image.

[1]M. Tico, E. Immonen, P. Ramo, P. Kuosmanen, J. Saarinen, "Fingerprint Recognition using Wavelet features", *IEEE International Symposium on Circuits and System(ISCAS)*, vol. 2, 2001, pp. 21-24.

Table VI: Results for Fingerprint Recognition

Feature Extraction by	***1*-NN (%)**	***3*-NN (%)**	**5-NN (%)**
Curvelets ($S = 3$, A = 32)	100	100	97.77
Wavelets (Symmlet 9, 5)	93.33	93.33	86.66
Wavelets (Symmlet 8)	93.33	93.33	82.22
Wavelets (Symmlet 6)	93.33	93.33	80.00
Wavelets (Symmlet 4)	93.33	93.33	82.22

VITA AUCTORIS

Tanaya Mandal was born in 1982 in Howrah, West Bengal, India. She graduated from Indian Institute of Engineering Science & Technology (Bengal Engineering College), Shibpore, India, with a Bachelors' Degree in Electronics and Telecommunication Engineering in June, 2005. She has worked as a software professional in PriceWaterhouseCoopers Ltd., Kolkata, India for one year. Presently Ms. Mandal is a pursuing her Master of Applied Science degree in the department of Electrical and Computer Engineering, University of Windsor, Ontario, Canada, with specialization in Biometrics/Pattern Recognition under the guidance of Dr. Jonathan Wu and expects to graduate in June 2008.