2010

Facial Expression Recognition Using Multiresolution Analysis

Ashirbani Saha
University of Windsor

FACIAL EXPRESSION RECOGNITION USING MULTIRESOLUTION ANALYSIS

by

Ashirbani Saha

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

2010

Facial Expression Recognition Using Multiresolution Analysis

by
Ashirbani Saha

APPROVED BY:

Dr. Jessica Chen, Outside Departmental Reader,
School of Computer Science

Dr. Narayan Kar, Departmental Reader,
Electrical and Computer Engineering

Dr. Jonathan Wu, Advisor,
Electrical and Computer Engineering

TBA, Chair of Defense

10 May 2010

Declaration of Previous Publications

This thesis includes 2 original papers that have been previously submitted/published for publication in peer reviewed conferences, as follows:

Chapter	Publication title/full citation	Publication status
Chapter 4	Ashirbani Saha and Q.M. Jonathan Wu, "Facial Expression Recognition Using Curvelet based Local Binary Patterns"; IEEE International Conference on Acoustics, Speech and Signal Processing, 2010 © IEEE, Reprinted with permission	Published
Chapter 5	Ashirbani Saha and Q.M. Jonathan Wu, "Curvelet Entropy for Facial Expression Recognition"; Pacific-Rim Conference on Multimedia, 2010	Submitted

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my thesis. I certify that the above material describes work completed during my registration as graduate student at the University of Windsor.

I declare that, to the best of my knowledge, my thesis does not infringe upon anyones 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.

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 or Institution.

Co-Authorship Declaration

I hereby declare that this thesis incorporates material that is result of joint research, as follows:

This thesis incorporates the outcome of a joint research under the supervision of professor Dr. Q. M. Jonathan Wu. The collaboration is covered in Chapter 4, 5 and 6 of the thesis. In all cases, the key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the author, and the contributions of co-author was primarily through the provision of proof reading and reviewing the research papers regarding the technical content.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-authors to include the above materials in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

Abstract

Facial expression recognition from images or videos attracts interest of research community owing to its applications in human-computer interaction and intelligent transportation systems. The expressions cause non-rigid motions of the face-muscles thereby changing the orientations of facial curves. Wavelets and Gabor wavelets have been used effectively for recognition of these oriented features. Although wavelets are the most popular multiresolution method, they have limited orientation-selectivity/ directionality. Gabor wavelets are highly directional but they are not multiresolution methods in the true sense of the term. Proposed work is an effort to apply directional multiresolution representations like curvelets and contourlets to explore the multiresolution space in multiple ways for extracting effective facial features. Extensive comparisons between different multiresolution transforms and state of the art methods are provided to demonstrate the promise of the work. The problem of drowsiness detection, a special case of expression recognition, is also addressed using a proposed feature extraction method.

to

The Almighty God,

my husband Dibyendu

and

my parents

Acknowledgements

I would like to express my deep-felt gratitude to my advisor, Dr. Jonathan Wu for his guidance, encouragement and support throughout her masters' studies. I would like to thank Dr. Narayan Kar and Dr. Jessica Chen for their precious suggestions and guidance. I would like to take this opportunity to thank Dr. Boubakeur Boufama, Dr. Dan Wu and Dr. Nihar Biswas for their guidance and help.

I would like to thank my friends and colleagues Rashid, Gaurav, Pankaj, Adeel, Aryaz and Thanh for their help and support. I thank Shaui and Fatih for their help. I thank my friends and well-wishers Tanaya, Shilpi, Sarmistha, Sanjib, Himadri, Rajashri and Saptarshi for their support and making the last two years a memorable one.

I thank my family for their love and support. I thank my husband specially for standing by me in all situations.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAM	Active Appearance Models
AU	Action Unit
DFB	Directional Filter Bank
DWT	Discrete Wavelet Transform
FACS	Facial Action Coding System
FT	Fourier Transform
HCI	Human-Computer Interaction
HMM	Hidden Markov Model
JAFFE	Japanese Female Facial Expression
LBP	Local Binary Pattern
LVQ	Learning Vector Quantization
LDA	Linear Discriminant Analysis
LP	Laplacian Pyramid
MLP	Multiple Layer Perceptron
MR	Multiresolution Representation
MRA	Multiresolution Analysis
PCA	Principal Component Analysis
PERCLOS	Percentage of Eye Closure

STFT Short Term Fourier Transform

SVM Support Vector Machines

Chapter 1

Introduction

This thesis focuses on the area of facial expression recognition with the help of a signal processing method called multiresolution analysis.

1.1 What is Facial Expression Recognition?

Facial expression recognition is the task of identifying mental activity, facial motion and facial feature deformation from still images, image sequences or videos, and classifying them into abstract classes based on the visual information only. Generally, this is possible because human facial gestures are similar. Internal feelings of humans are often reflected spontaneously on their faces thereby changing the fiducial appearances and dynamics, and making the face '*index of mind*'. Thus, expression recognition helps to interpret mental states and differentiate between facial gestures. Figure 1.1 shows a neutral face and six expressions posed by a subject of the popular Japanese Female Facial Expression (JAFFE) database [33].

Figure 1.1: Different facial expressions [33]

1.1.1 Natural Similarity of Facial Expressions

The question of recognition or identification arises due to the presence of similarity. The emotions expressed by humans deform their faces. There exists natural similarity of facial expressions despite the presence of diversities in ethnicity, age or gender. This fact was first demonstrated by Charles Darwin [11]. After nearly a span of 100 years, Ekman and Friesen [18] postulated six expressions namely happiness, sadness, surprise, anger, disgust and fear (also known as *basic emotions*) each having a special and distinguishable nature of its own. This similarity in human facial expressions is exploited by all facial expression recognition systems.

1.1.2 Facial Expression Identification by Coders

In spite of apparent similarity in expressions, facial expression recognition is a tough task due to the location, inherent variations in intensity and dynamics of facial expressions. Moreover, the occurrence of expressions and their durations vary for different people. The expressions do not appear all of a sudden; the facial changes take place gradually (though very fast). Thus, any expression has three phases: *onset* or attack, *peak* or when the expression sustains on the face, and finally $offset$ or relaxation of the fiducial muscles into the original state. Often human expressions do not present single unmixed mental state. In view of this, there is a need for preparing ground truth for expression data. The ground truths are generated in two ways, namely (a) Judgement-based approaches and (b) Sign-based approaches. The judgement-based approaches, as the name suggests, depend on the decision of certified human coders where the decision of the majority is considered as the ground truth of the image. The images are classified into the basic six emotional states [18] in this case. In sign-based approaches, the facial motion and deformation is assigned to classes each of which serves as a very basic unit of a subtle facial muscular movement. Thus, a complete dictionary of basic facial movements is formed, and it is called Facial Action Coding

System (FACS). Each fundamental facial movement enlisted in FACS is named as an Action Unit (AU). The abstract classes in which the expressions are categorized by FACS are facial movements along with their locations and intensities. FACS was proposed by Ekman and Friesen [42]. The facial expression research is therefore directed towards developing robust algorithms that can predict the AUs or basic emotions accurately using visual data and proceed towards a completely automated expression recognition process.

1.2 Facial Expression Recognition as a Research Topic in Computer Vision

As correctly put by an old saying '*Necessity is the mother of invention*', the driving force behind any research field is its potential applications or its ability to solve problems. It has been the aim of all computer vision and image processing algorithms to make use of visual information as much as possible. Facial expression recognition in computer vision started with the same aim. With the advent of cheap computational power and improvement in related fields of face processing, facial expression recognition gained much inertia. The activeness of this research field enhanced owing to its applications in Human-Computer Interaction (HCI), where human voice, visual appearance of people as well as sensations of sight and touch (which are also called *modes*) are used simultaneously. Also, as social psychology claims, facial expressions help to co-ordinate conversation. According to Mehrabian [37], when a spoken message is delivered with visual information as well, spoken words contribute 7%, vocal tone provide 38% but above all, facial expressions generate 55% of the effect of message. Thus, among all of the modes in HCI, facial expression is the most important one. This has necessitated the research of facial expression recognition. Besides this area, robot vision, virtual reality and facial animations require facial expression anal-

ysis. Intelligent transportation systems, which are targeted to eliminate human lapses in vehicle control, require facial expression recognition to make decisions. Research has been carried out on two genres of expressions (a) recognition of AUs (b) recognition of basic emotions. The recognition of action units emphasizes on capturing local muscular deformation of facial muscles. On the other hand, basic emotion recognition aims to track mental states.

1.3 Significance of Expression Databases

Generally, recognition algorithms are checked against specific image or video databases which contain images and their corresponding ground truths. Databases provide an uniform platform for algorithm testing. The databases gather images under controlled situations as well as incorporate diversity in their content in terms of different subjects, illuminations, ages, gender ethnicities, textures, in general. Any algorithm, performing robustly in the databases, can be checked for evaluation in more critical situations and improved subsequently. The databases have been generated by earlier researchers for their work and they have generously made them publicly available for further usage by other researchers. Many face processing databases are available currently, and among them the JAFFE database and the Cohn-Kanade database [27] are the most used ones for facial expression analysis.

1.4 Problem Statement

In this thesis, images have been decomposed by Multiresolution Analysis (MRA) methods and the Multiresolution Representations (MRs) of images have been used to analyze facial expressions, specifically, basic prototype emotions from still images. The aim has been to study and compare the application of MRA methods for expression recognition in the JAFFE and Cohn-Kanade databases. Two novel facial ex-

pression recognition methods based on multiresolution methods have been proposed. Also, a special case of expression analysis (applicable in intelligent transportation systems) called *drowsiness detection* from still images has been studied using MRs. Since, there are no databases publicly available for drowsiness, a database has also been developed.

1.5 Thesis Outline

A brief outline of the rest of the thesis is discussed here.

Chapter 2 discusses the earlier works in the field of facial expression recognition. The different sub-areas of expression recognition are presented along with the earlier researches carried out. Also, the chapter sketches an outline of the different types of approaches existing in the field.

Chapter 3 presents the fundamental theory of the MRA methods. The chapter starts with the evolution of MRA from the fourier transform. Abiding by their respective times of arrival in literature, Gabor wavelets, wavelets, curvelets and contourlets are discussed.

Chapter 4 comprises of a study of facial expression recognition using the novel idea of curvelet based Local Binary Patterns (LBPs). The comparison of the method with some of the state of the art methods is also presented here.

Chapter 5 presents the novel use of multiresolution entropy at selected points for facial expression recognition. A comparative study on facial expression recognition using different multiresolution based entropies is carried out.

Chapter 6 studies drowsiness detection using MRs of faces. Two methods involving different classifiers have been proposed for drowsiness detection. It also discusses the generation of drowsiness database for the purpose of this work.

Chapter 7 concludes the thesis by discussing the contributions of the work and highlights the scopes of the future work based on the thesis.

Chapter 2

Literature Survey

The first step towards facial expression analysis using images was taken by Suwa, Sugie and Fujimora [50]. But the pioneering work of Mase and Pentland [36] was the one to inspire other researchers in this field. Surveys of the work done in this field till the last decade is found in the works of Fasel et. al and Pantic [21, 44]. A glance, from now, at the previous works broadly divides the fields into two parts; (a) Static-image based recognition (b) Image sequence or video based recognition. Static image based researches rely on peak or apex of the expressions and process still images, and hence, they miss subtle details of expressions. But, these methods are faster and simpler compared to their image sequence or video based counterparts that can capture the dynamics of expressions. In spite of this dichotomy, all facial expression recognition approaches abide by a generic framework.

2.1 The Generic Framework for Expression Recognition

The generic framework for all expression recognition systems is presented in Fig. 2.1. The basic steps for expression recognition and the existing works on them in the literature are discussed in the following sub-sections.

2.1.1 Face Acquisition

The aim of face acquisition is to locate faces in complex scenes with cluttered background. For automatic expression analysis systems, the perfect detection of face

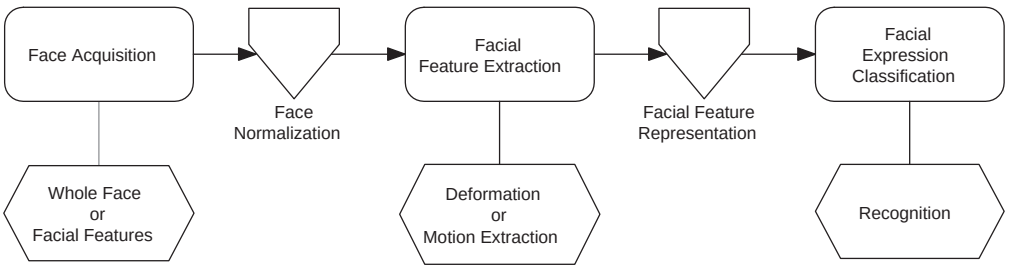


Figure 2.1: Generic framework for facial expression recognition systems

is very critical. Some approaches [19, 24] use exact location of faces whereas others [2, 29] can work with the coarse location of faces. Essa and Pentland [19] used modular eigenspace method for exact face detection. The Principal Component Analysis (PCA) coefficients of facial images were used to detect faces from still images as well as image sequences. Hong et. al [24] used the PearsonSpotter system [49] for head tracking and thereafter stereo disparity, skin colour detection and convex region detector for figuring out the exact face locations. Methods like Active Appearance Models (AAM) [29] and local motion model [2] can work with a coarse location of faces. Actually, these methods start with an approximate location of faces and adapt with the exact location of faces by minimizing some mathematical distance functions. Hence, essentially the exact face is processed for recognizing the expressions. The popular face detector used in recent research works [1, 46] is the Viola-Jones face detector [56] which uses boosted haar features to detect faces.

2.1.2 Face Normalization

Face normalization is the process of rectifying face appearance changes caused by in-plane rotation, out-of-plane rotation and illumination variations. This processing is necessary to bring uniformity in the faces, and facilitates the generation of reliable facial features. In-plane rotations occur due to the rotation of the face parallel to

the camera plane. Hence, the image of the face is captured fully but the face is not oriented vertically. In-plane rotations are comparatively easier to handle and two-dimensional spatial transforms are used for adjusting them. A very popular normalization technique for in-plane rotation is Tian's method [53] to produce facial images that have fixed distance between the eyes. Out-of-plane rotations occur when the face rotates making an angle with the camera plane. Hence, the images may not have a full view of the face. If very little part of the face is captured in the image, then expression analysis cannot be continued. However, out-of-plane rotations can be dealt with warping techniques as shown in [19]. Also Pantic and Rothkrantz [43] used a point based model for side view of face. For illumination variations, Gabor wavelets have been used in [22, 33] as the cosine part of Gabor filter can remove the average illumination of the image.

2.1.3 Facial Feature Representation

Generally, in the field of image processing or computer vision, feature representation is the process of generating useful information from the images/image sequences/videos, such that, effective use of the information can be robust enough to categorize the images into correct abstract classes relevant to the context. A digitized image is a two-dimensional (three-dimensional for a colour image) matrix; image sequences or videos are three-dimensional (four-dimensional for colour) matrices. The basic idea is to apply mathematical methods to generate effective information from them. Figure 2.2 shows different genres in which the methods are broadly classified. The first dichotomy in face processing arises from the proportion of face that is considered for processing. There exist local methods [16, 43] which use, mainly, the portion of eyes and mouth for processing. The entire face or points from different regions of the face are used in holistic methods [2, 24, 46]. There also exist approaches that use both local and global features and hence are termed as hybrid methods [19, 61].

Irrespective of the face being processed globally or locally, video or image sequence

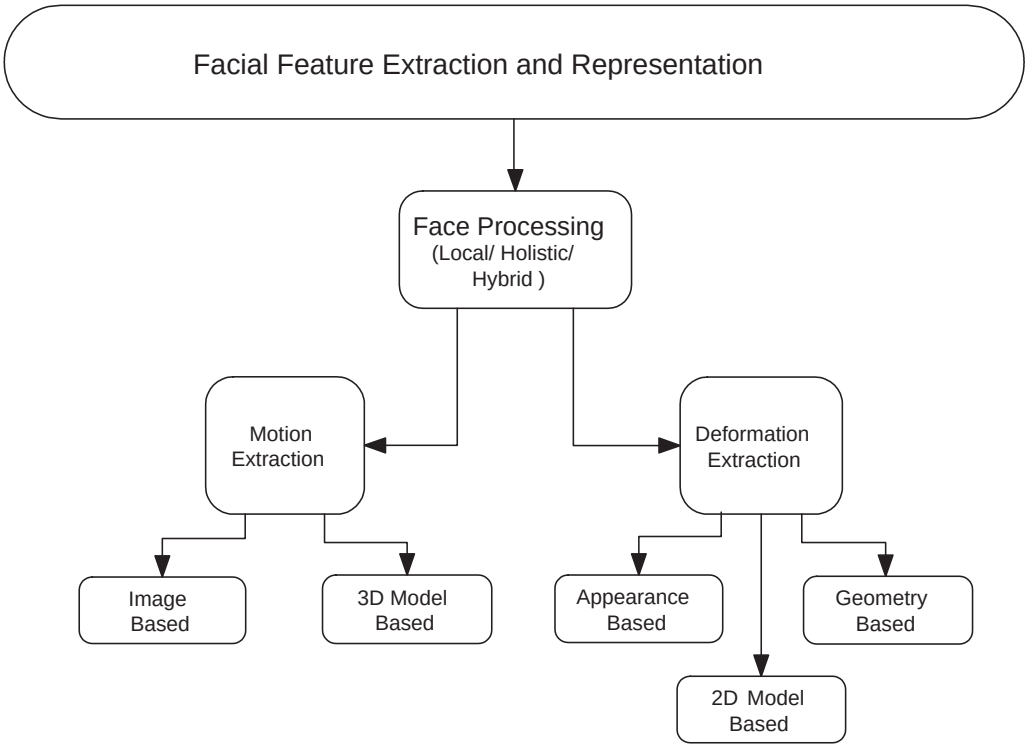


Figure 2.2: Different types of facial feature extraction approaches

based methods extract motion and image based methods extract deformation. Motion based methods [17, 19, 36, 54, 59] aim to obtain the subtle movement of facial parts in between frames. Mostly dense optical flow [31, 59] and 3D deformable models [17, 19] have been used for motion extraction. Essa and Pentland [19] used Kalman filter for reliable tracking followed by 3D motion and muscle models for feature extraction. However, the model based method undergoes through the tedious process of model generation and they are also computationally expensive. Mase and Pentland used regional optical flow [36] which was able to capture the motion of 44 facial muscles. Their approach used a window for each muscle and each window had an axis by which the optical flow was estimated in eight directions. Though the optical flow methods are vulnerable to noise, their sensitivity in motion discontinuity inspired their use. It is found that motion extraction methods may be image based or 3D

model based but Essa and Pentland [20] used the information obtained from optical flow to map fiducial muscle movements. Recently, dynamic Haar like features were used by Yang et. al [60] to extract features from videos or image sequences. Dornaika and Davione [15] used particle filter for recovering facial dynamics from videos.

Deformation extraction can be classified into geometry based methods, appearance based [10, 32, 61] methods and two-dimensional model based methods [9, 16, 29]. Geometry based methods form features that depend on the shape and location of facial features. It has been showed by Valstar and Pantic [55], that appearance based methods perform better than geometry based methods. In appearance based methods, Gabor wavelets [1, 61] have been used to form features. Both of these approaches select specific points from facial images and use the Gabor wavelet coefficients at those points for further processing. Recently, LBPs have also been used to form effective features for expression recognition [23, 46] and it is proved that the use of LBPs reduces the complexity of the algorithms when compared to Gabor wavelets. These methods use holistic face processing. Recent work on static images focus on reducing time [45, 46] as well as on the use of the automated system [45] for facial expression analysis. The two-dimensional model based methods are mainly dependent on AAM which generates a statistical model of shape variation using 122 fiducial points and PCA. A multiple multivariate regression analysis is carried out for learning in AAM. The learned model is used to classify the images. Generally, the features are represented in form of vectors called $feature\ vectors$ for classification.

2.2 Facial Expression Classification

Facial expression classification is the final step, where, a query image is classified into its object class. After the earlier activities of face acquisition, normalization, facial feature extraction and representation are performed, the facial features are used for classifying the images or expressions to the abstract classes. The classification is done

by a mathematical abstraction called *classifier*. The classifier is taught about the features in its training phase so that, in the testing phase, the classifier can allot the feature vector(s) from a query image to its proper class. Generally, features are extracted from all the images/videos that are intended for training to form a feature set. This feature set is used to train a classifier. After the classifier is trained, the query image is used to test the classifier results. Different types of classifiers have been used in existing works. Nearest neighbour classifiers have been used in the works of Shan et. al. and Liao et. al. [46, 30] for classifying feature vectors. Support Vector Machines (SVM) have been used as the classifier in some works [46, 38]. Several works have used neural networks; Multiple Layer Perceptron (MLP) has been used by Zhang [61], Learning Vector Quantization (LVQ) have been applied by Bashyal and Vengamoorthy [1]. The recent researches have used adaboost [46, 60] for enhanced performance. These are mainly used as spatial classifiers. Hidden Markov Models (HMMs) [41] and *Recurrent Neural Networks* [28] have been used as spatio-temporal classifiers.

Chapter 3

Multiresolution Analysis Methods

The word *multiresolution* points to the meaning- multiple resolutions. Very simply, MRA methods are concerned about gathering information using the representation and analysis of signals or functions in more than one resolution. Generally, MRA methods are mathematical frameworks for decomposing a signal into its relevant constituents. An image is considered as a two-dimensional signal and can therefore be generalized as a part of two-dimensional vector space V which is a set closed under finite vector addition and multiplication. MRA methods consider the image as the vector space in the highest available resolution such that it contains all vector spaces of the next lower resolution. The next lower resolution contains all vector spaces of the further lower resolutions. Thus subspace nesting is noticed and that continues until the two-dimensional image reduces to a two-dimensional point (basically to a pixel). MRA methods offer to find effective vector spaces that can synthesize the original signal from the decomposed parts. Vector spaces are characterized by a subset (which may not be unique) $v_1,\ v_2, ..., \ v_n$ of vectors in V such that these vectors are linearly independent and span V (which means every vector v in V can be generated by a linear combination of the elements of the subset). A subset of this kind is called a basis of the vector space. Thus, an MRA method needs to find effective basis for signal decomposition and reconstruction. Physically, MRA method provide a simple hierarchical framework for analyzing the properties of a signal by measuring it and represents it effectively so that the signal can be reconstructed. A key aspect of this framework is its hierarchical structure. At different resolutions, the details of a signal, primarily, characterize different physical aspects of a signal. At

coarse resolution, these details correspond to the larger overall aspects, the context of a signal. At fine resolution, these details correspond to its distinguishing features. Such a course-to-fine strategy is typical in identifying a signal. This aspect of MRA is exploited in this work.

The inspiration of multiresolution analysis came from the shortcomings of Fourier Transform [3] which is used to generate a frequency domain description of a signal using sine and cosine functions as the basis elements. Given a time limited signal, Fourier Transform (FT) can provide the frequency components(with value and phase) required to synthesize the signal within its time limits. However, FT of a signal stretches in frequency domain as the time limits of the signal decrease making it impossible for the transform to conclude which frequency components are responsible for the generation of an arbitrarily small time-limited part of the signal. This phenomenon is called *Uncertainty Principal* and MRA methods try to solve it as much as possible. The rest of the chapter presents the evolution of MRA methods.

3.1 From Fourier Transform to Gabor Wavelets

The FT of $g(t)$, a time varying signal is mathematically represented as follows

$$G(w) = \int_{-\infty}^{\infty} g(t)e^{-jwt}\,dt \tag{3.1}$$

where $G(w)$ is the description of $g(t)$ in frequency domain. Here, t represents time and w represents frequency. $G(w)$ is called the forward FT of $g(t)$. There also exists a reverse FT(shown in eqn. 3.2) by which $g(t)$ can be recovered by $G(w)$.

$$g(t) = \int_{-\infty}^{\infty} G(w)e^{jwt}\,dw \tag{3.2}$$

Generally, t can represent any quantity. It has been used as time here for the simplicity of the discussions. From the equations 3.1 and 3.2, it is evident that time domain signal $g(t)$ does not have frequency domain information and the frequency

domain signal $G(w)$ can only say that it has specific frequency components but cannot localize them in time domain. In other words, $g(t)$ has perfect time resolution but no frequency resolution and $G(w)$ has frequency resolution but no time resolution. Therefore fourier transform is suitable for those signals where same frequency components are present at all instances of time. These signals are called stationary signals. For non-stationary signals, the idea of Short Term Fourier Transform (STFT) came into being. The assumption is that a non-stationary signal can be divided into small stationary parts using a kernel or window function. The product of the signal and kernel function localized at a specific instant of time can be subjected to FT as shown in eqn. 3.3 to obtain the STFT or *windowed* FT of the signal.

$$G(\tau, w) = \int_{-\infty}^{\infty} g(t)p(t-\tau)e^{jwt}\,dw \tag{3.3}$$

where $p(t-\tau)$ represents the kernel function with its peak present at the instant $t = \tau$. FT and STFT are used mostly for acoustic signal processing. Therefore, from eqn. 3.3, it is clear that G is a function of two variables– time (τ) as well as frequency (w). Hence, STFT provides a time-frequency description of signals. But, there exists a *quantum principle* by which the conjoint time-frequency domain cannot reach below a certain value. Hence, a tradeoff between time resolution and frequency resolution was needed. D. Gabor discovered that gaussian modulated complex exponentials could provide the best tradeoff. This concept is used by the name of Gabor functions or transform. Two-dimensional (typically, planar space used in image processing applications) gabor functions happen to be a special case of STFT where local bandpass filters are used to achieve the theoretical limit between two-dimensional spatial and two-dimensional fourier domains. The Gabor function is formed by a Gaussian function with its own characteristic mean and variance, and, a complex exponential (which represents a wave) of characteristic phase and frequency. The Gabor function is shown as

$$\Psi(\overrightarrow{k}, \overrightarrow{x}) = \frac{\|\overrightarrow{k}\|^2}{\sigma^2} exp(-\frac{\|\overrightarrow{k}\|^2 \|\overrightarrow{x}\|^2}{2\sigma^2})[exp(i\overrightarrow{k}.\overrightarrow{x}) - exp(-\frac{\sigma^2}{2})] \tag{3.4}$$

where $\overrightarrow{k}$ is called the characteristic wave frequency, $\overrightarrow{x}$ represents the position vector of a point in the $X-Y$ plane and σ is the standard deviation of the Gaussian function. $\overrightarrow{k}$ has two parameters, scale (k_υ) and orientation (θ).

$$\overrightarrow{k} = \begin{pmatrix} k_\upsilon cos\theta \\ k_\upsilon sin\theta \end{pmatrix} \tag{3.5}$$

Generally, the gaussian mean is taken as zero; k_υ and θ are changed to obtain the Gabor filter banks. Since, different variations are obtained by changing a mother function in scale and orientations, they are also called Gabor wavelets keeping similarity with the wavelets which are discussed in the following section. In applications involving facial images, the images are convolved with the Gabor filter banks to obtain Gabor coefficients at different scales and orientations. The problem that remains in Gabor transform is the use of fixed window size. In wavelets, the modification made is the use of varying window size depending on the frequency which essentially means large windows are used for low frequencies and small windows are used for high frequencies.

3.2 Wavelet Transform

The name wavelet means a part of wave or a small wave. Essentially wavelets are time-limited signals and they are localized in frequency [34]. Wavelets serve as the basis elements in wavelet transform. Their transient nature makes them totally different from the basis of fourier transform. The wavelets are smooth and have zero integral and this implies that they have no high or low frequencies. Wavelets are translates and dilates of one function called mother wavelet φ. Therefore, one dimensional wavelets can be mathematically represented as

$$\varphi_{s,d}(t) = \frac{1}{\sqrt{s}}\varphi(\frac{x-d}{s}) \tag{3.6}$$

where s denotes scale (dilation) and d denotes position (translation). The continuous wavelet transform $W_g(s,d)$ of a signal $g(t)$ can be obtained as

$$W_g(s,d) = \int g(t)\varphi_s(t)\,dt \tag{3.7}$$

In the discrete domain the wavelet transform is referred as Discrete Wavelet Transform (DWT), and 1D DWT of $g(t)$ is calculated as

$$W_\Phi(s_0,d) = \frac{1}{\sqrt{M}}\Sigma g(t)\Phi_{s_0,d}(t) \tag{3.8}$$

$$W_\varphi(s,d) = \frac{1}{\sqrt{M}}\Sigma g(t)\varphi_{s,d}(t) \tag{3.9}$$

Here Φ is called the scaling function which is similar to that of φ. $W_\Phi(s_0,d)$ gives the approximate or low frequency part of the signal and $W_\varphi(s,d)$ gives the high frequency part. s_0 is chosen as 1 and M is always in the power of 2. Thus,

$$M = 2^p, \tag{3.10}$$

where p is a positive integer. The scales and positions are chosen on powers of two which are called dyadic scale and positions such that $s = 1, 2, 3, ..., p-1$ and $d = 0, 2^0, 2^1, ..., 2^{p-1}$. Since, wavelets work on different scales, they are also called multiscale transforms. The idea of multiscale transforms, though, was first generated from the idea of Laplacian Pyramid (LP) [4] which decompose a signal at multiple scales with the aid of filter banks. Depending on the type of wavelets, there exists a family of wavelets that comprises of Haar, Daubechies, Morlets, Symmlets, Coiflets, Biorthogonal etc. Each of them have a characteristic nature of their own. By the property of wavelets, a one level wavelet decomposition of two-dimensional signal gives four subbands containing approximate, horizontal, vertical and diagonal information of the signal. The approximate subband can be further decomposed to approximate, horizontal, vertical and diagonal subbands making the overall wavelet decomposition level as 2 and reduce the size by 2^2. Thus, more levels implies more scales and positions of the signals are explored. This makes the transform work for just $0°, 45° and 90°$

orientations (of edges). The elements of the subbands are called wavelet coefficients. Also, many coefficients of wavelets are required to represent a discontinuity. These drawbacks inspired researchers to develop new MRA methods like contourlets and curvelets which can work in more orientations and obtain a sparse representation of signals. Since these transforms work in different orientations, they are also called directional.

3.3 Curvelet Transform

Curvelet transform was introduced by Candès and Donoho. Mathematical details of the transform can be found in [14]. The basis elements of the transform are anisotropic (unlike the dyadic wavelets) and they abide by the parabolic scaling law of $width \sim length^2$. This property takes a major role in obtaining sparse representation of smooth functions and straight edges by the curvelets. Figure 3.1 shows a comparison of edge representation by curvelets and wavelets. For obtaining curvelet transform of an

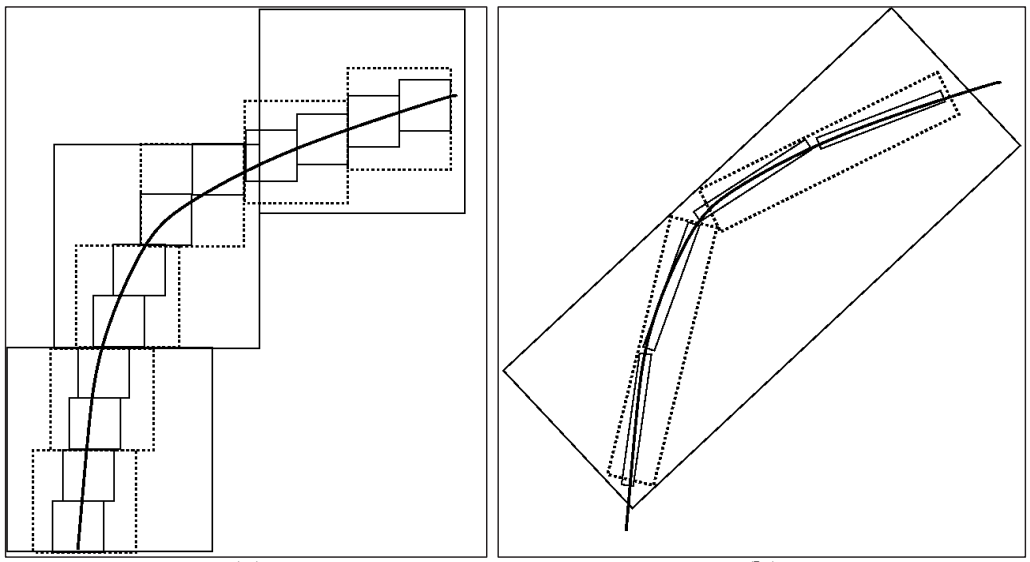


Figure 3.1: Edge representation using (a) wavelets and (b) curvelets [47]

image, at first, the image is represented in scale-space domain. Then the frequency plane obtained is divided into dyadic coronae and each corona is partitioned into angular wedges which abide by the parabolic aspect ratio. Hence, orientation is taken into account in the scale-space description of the image. Thus, curvelets provide coefficients by measuring information about an object at specified scales, locations as well as orientations. As explained by Candes, curvelets can be thought of as obtained by applying parabolic dilations, rotations and translations to a specifically shaped function Φ which is indexed using scale $s(0 < s < 1)$, location l and orientation β as

$$\Phi_{s,l,\beta}(x) = s^{-\frac{3}{4}}\Phi(D_a R_\beta(x - l)) \tag{3.11}$$

with

$$D_s = \begin{pmatrix} \frac{1}{s} & 0 \\ 0 & \frac{1}{\sqrt{s}} \end{pmatrix}$$

where D_s is a parabolic scaling matrix, R_β is a rotation by β radians.

There are some digital implementations of curvelets and *curvelets via wrapping* [6] (also called second generation of curvelets) are used for this work. For a 2D function $g[x_1, x_2]$ with $0 < x_1, x_2 < n$ (a specific length), the curvelets via wrapping can be performed according to the following steps[6]

1. 2D Fast Fourier Transform (FFT) for $g[x_1, x_2] \rightarrow \hat{g}[w_1, w_2]$ is computed.

2. $G[w_1, w_2]$ is divided into dyadic subbands using a scale window V_i, i representing i^{th} scale.

3. Each subband is separated into angular wedges using angular windows $V_{i,k}$ with k representing k^{th} wedge.

4. The product $\tilde{P}_{i,k}[w_1, w_2]G[w_1, w_2]$ is calculated where $\tilde{P}_{i,k}[w_1, w_2]$ is a discrete localizing window.

5. The product is wrapped inside a rectangle R of size $H_{1,i} \times W_{2,i}$ (in east-west) or $H_{2,i} \times W_{1,i}$ (in north-south) around the origin to obtain $\tilde{G}_{i,k}[w_1, w_2] = R(\tilde{U}_{j,l}G)[w_1, w_2]$. Here, $H_{1,i} \sim 2^i$ and $H_{2,i} \sim 2^{\frac{i}{2}}$ and are constants.

6. The curvelet coefficients at scale i and orientation k are calculated from the 2D inverse FFT of each $\tilde{G}_{i,k}$.

Curvelet transform is applied in different scales and orientations for this work.

3.4 Contourlet Transform

Contourlet transform [13] decomposes an image into several directional subbands at multiple scales and obtains a sparse image expansion by first applying a multiscale transform followed by local directional transform. For this purpose, the edges or point discontinuities are detected by a LP. A Directional Filter Bank (DFB) is used thereafter for linking those points into linear structures. The overall work is achieved by *pyramidal directional filter bank* [12] and finally a multiscale image expansion using elementary images like *contour segments* are obtained. Therefore, the name contourlet transform is given to the process. The DFB used here is different from the traditional one which modulates the signal during frequency partitions. This is because traditional DFB was developed for high frequency and low frequency was not handled properly. During the implementation, bandpass images obtained from an LP are fed to a DFB to retrieve the directional information in the first level. This scheme is iterated on the approximate image for the next level. Thus, the formation of *contourlet filter bank* takes place in the guise of a doubly iterated filter bank. This filter bank is shown in Fig. 3.2.

Contourlet transform offers multiresolution and directional decomposition for images and it allows for a different number of directions at each scale. Like curvelets, contourlet transform satisfies the anisotropy scaling law. The contourlet transform is

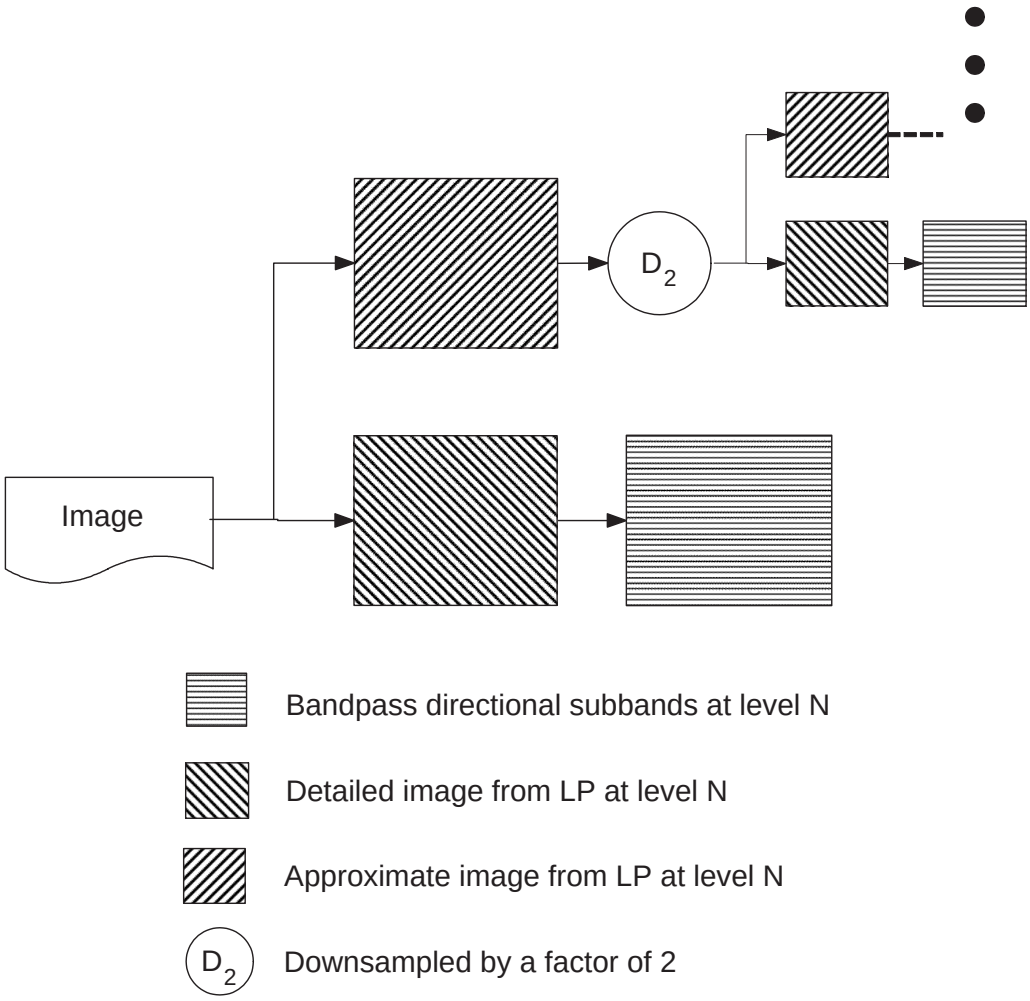


Figure 3.2: Contourlet filter bank

almost critically sampled and much less redundant compared to the curvelets. Contourlet transform has been used in various combinations of scale and orientations for carrying experiments in this work.

Chapter 4

Curvelet based Local Binary Patterns for Facial Expression Recognition

4.1 Introduction

This chapter studies the use of curvelet based LBP for the facial expression recognition. Curvelet based LBP are used to extract features from the face. Though LBP [46, 30] has been used earlier for facial expression, the use of LBP on multiresolution space has not been studied earlier. The motivation of this work is derived from the recent use of curvelets on the problem of face recognition [35]. Generally, face recognition and facial expression recognition are dual problems. Face recognition is made difficult by variety of expressions and expression recognition gets tougher due to the faces varying in age, gender, and ethnicity. The method presented in [35] applied curvelet coefficients to form features for representing the entire face. In order to classify the facial expressions, the local facial information needs to be stored. To obtain the local description of the expressions, local binary patterns (LBPs) are computed using selected sub-bands of image pre-processed by curvelet transform. The approach is non-cross-validated and has been compared with LBP and Gabor wavlet based methods.

4.2 Local Binary Patterns

The LBP operator, introduced by T. Ojala et al. [39], is a method used in texture description and analysis. Using the basic LBP operator, each pixel of an image is

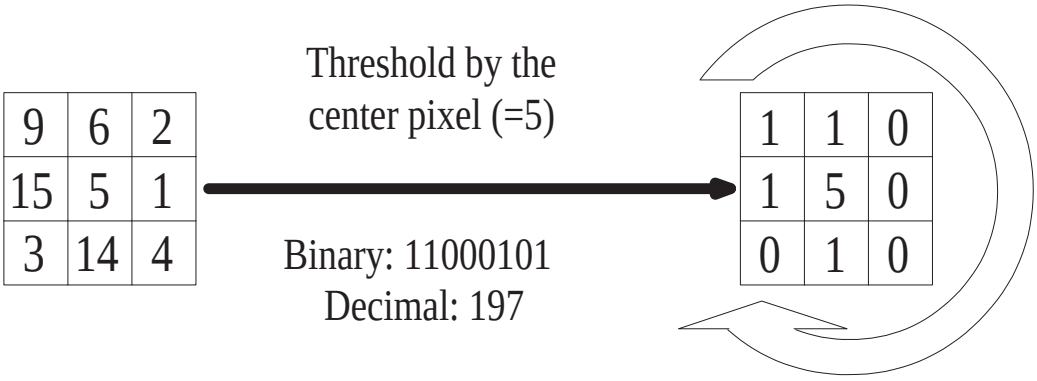


Figure 4.1: Basic LBP operator

labeled by a binary (0 and 1) representation of its 3×3 neighbourhood taking the center value as threshold as depicted in Fig. 4.1. Then the histogram of the labels can be used as a texture descriptor. A histogram of the labeled image $f_l(x, y)$ can be defined as

$$H_i = \sum_{x,y} I\{f_l(x, y) = i\}, i = 0, 1, 2, ..., n - 1 \tag{4.1}$$

in which n is the number of different labels produced by the LBP operator and

$$I\{A\} = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{if } A \text{ is false} \end{cases} \tag{4.2}$$

This histogram has information about the distribution of the local micropatterns, such as edges, spots and flat areas of the entire image [39]. The number of labels clearly depends on the number of neighbours used in the LBP operator. For N neighbours, number of labels will be maximum 2^N. The operator was modified to use neighbourhoods of different sizes in [40]. In the modified operator, the neighbourhood of any pixel was defined as a circular region of radius R. Using bi-linear interpolation, a number of P(depending on R) pixels can be sampled on the circumference of the circular region using the center pixel as threshold. The LBP is therefore represented by LBP(P, R). If the coordinates of the center pixel are (0,0), the coordinates of the pth neighbour are given by $(-R\sin(\frac{2\pi p}{P}), R\cos(\frac{2\pi p}{P}))$. Since the labels are formed using

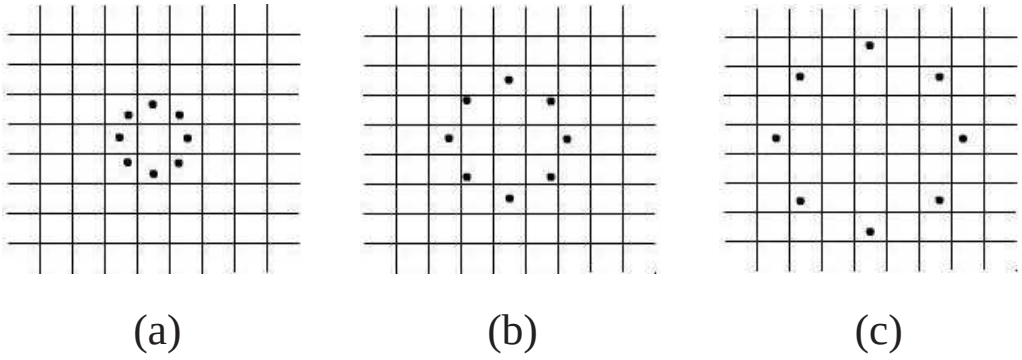


Figure 4.2: LBP of different radii and neighbourhoods (a)R=1, N=8 (b)R=2, N=8, (c)R=3,N=8

8 neighbours, 2^8 labels are available, starting from 0 to 255. LBP operators shown in Fig. 4.2 have been used in this expression recognition framework for experiments.

4.3 Proposed Method

The proposed method is shown in Fig. 4.3. The images are first cropped to extract

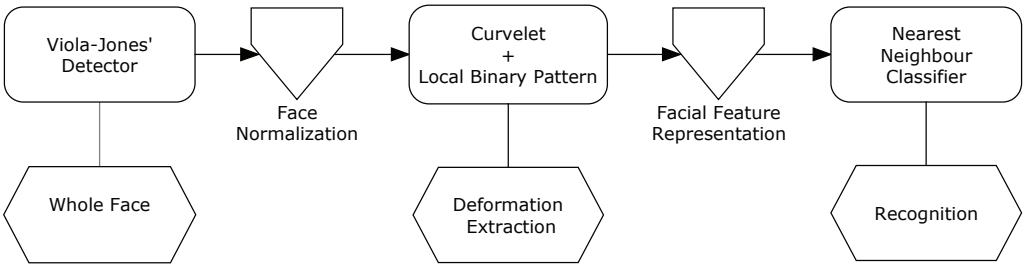


Figure 4.3: Proposed method using curvelet based LBP

the face of the subject using Viola-Jones face detector. LBP is independent of illumination changes whereas curvelet is not independent of illumination changes. Hence, normalization is applied. Histogram equalization is then applied to increase the contrast. After that, curvelet transform is applied on the images at a specific scale and

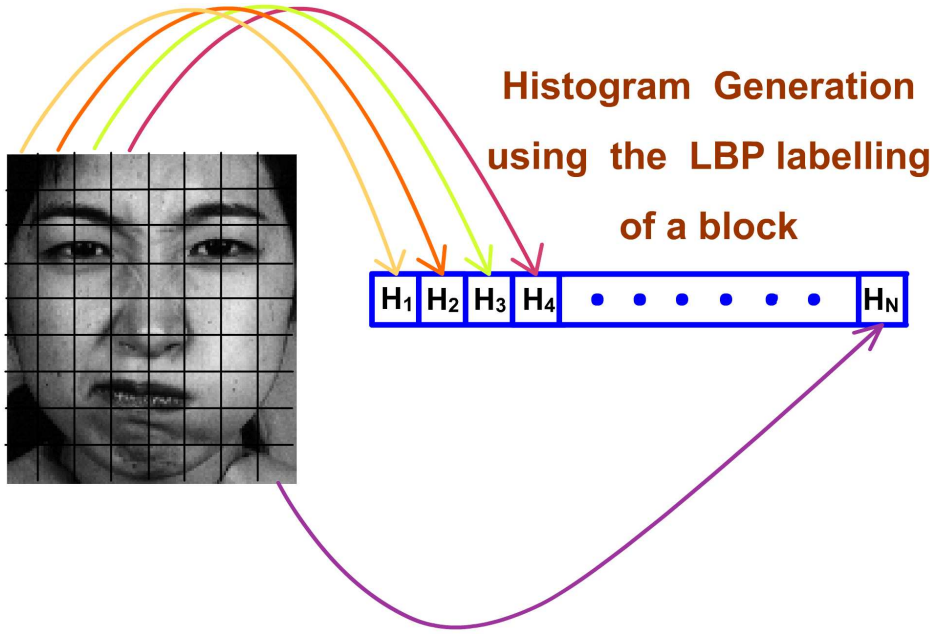


Figure 4.4: Feature generation using curvelet based LBP

orientation. The approximate sub-band obtained from the transform is resized to $a \times b$ and divided into k regions each of size $m \times n$ pixels. From each of these k regions, the LBP histogram of 255 labels is calculated. The histograms from successive regions are concatenated to form the feature set for a particular image. Mathematically, k regions $G_1, G_2, ..., G_k$ are available and each element of the feature vector can be expressed as

$$H_{i,j} = \sum_{x,y} I\{f_l(x,y) = i\} I\{(x,y) \epsilon R_j)\}. \tag{4.3}$$

Here, i=[0, 255] and j=[1, k]. Figure 4.4 shows the feature extraction process from the approximate subband of the facial image in curvelet domain. So, effectively from the approximate sub-band, three levels of information can be obtained. The LBP values have information about the coefficients. Histogram obtained from the LBP values over a region contain the information in a regional level. Finally, all the regional histograms are successively concatenated to obtain a holistic description of

the sub-band. Therefore, i^{th} feature vector x_i of length $k\times255$ is obtained. Based on the class labels y of the images, feature vectors of same class label are grouped to form the training set X^c for a particular class of expressions. Thus,

$$X^c = \{x_1^c, x_2^c, x_3^c, ..., x_n^c\} \tag{4.4}$$

where n is the number of training images available for the corresponding class. The representative feature set M^c of the class c is the cluster center of X^c and is calculated as

$$M^c = \frac{1}{n}\sum_{i=1}^{n} x_i^c \ . \tag{4.5}$$

A simple nearest neighbour classifier using the Chi-square metric (4.6) is used for classification.

$$\chi^2(S, M^c) = \sum_{i=1}^{N} \frac{(S_i - M_i^c)^2}{(S_i + M_i^c)} \tag{4.6}$$

Here, S is the feature vector of length N extracted from the test image.

4.4 Experimental Results using Curvelets

The experiments are carried out on JAFFE and Cohn-Kanade databases. The details of the experiments are discussed in sub-sections 4.4.1 and 4.4.2.

4.4.1 JAFFE Database

The database consists of total 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral, frontal view) posed by 10 Japanese female models. They are pre-processed as described in section 4.3. The images are randomly divided into five sets of roughly equal images. Five rounds of testing are carried out and at each time, a different combination of four sets are used for training and the remaining set is used for testing. This random division is carried out 3 times. The recognition rate is the percentage accuracy of the proposed method in classifying expressions. The recognition rates given for this database in the experimental results are therefore the averages

Table 4.1: Recognition rates obtained by different curvelet (scale, orientation) + LBP (neighbourhood, radius) combinations on JAFFE database

Combination Details	Recognition Rates (%)
curvelet(3,16) + LBP(8,3)	90.99
curvelet(3,16) + LBP(8,2)	93.69
curvelet(3,16) + LBP(8,1)	90.86
curvelet(3,8) + LBP(8,3)	90.11
curvelet(3,8) + LBP(8,2)	91.98
curvelet(3,8) + LBP(8,1)	89.07
curvelet(2,8) + LBP(8,3)	83.41
curvelet(2,8) + LBP(8,2)	91.75
curvelet(2,8) + LBP(8,1)	89.74

of the recognition rates of all tests. The recognition rates for various combinations of curvelet (Scale, Orientation) and LBP (Neighbourhood, Radius) are discussed in Table 4.1. Since the combination of curvelet (3,16) and LBP (8,2) gives best results, the confusion matrix shown in Table 4.2 is calculated using this combination. Also, if the curvelet scale and orientation parameters are kept fixed, the LBP(8,2) gives the best results among other LBP parameters. The confusion matrix shows the proportion in percentage, any expression shown in a row is falsely detected as another expression in the column. The ideal confusion matrix will be a diagonal matrix the expression From the confusion matrix of JAFFE database, it is found that happiness is least confused with other expressions. After happiness, disgust is detected with least error. Though sadness is confused with fear only, it suffers the least accuracy among all expressions. Sadness and fear are found most difficult to detect and these results are similar with the earlier results published.

Table 4.2: Confusion matrix (%) obtained by curvelet (3, 16) and LBP (8, 2) on JAFFE database

Given Label \ Classified Label	Happy	Sad	Surprise	Anger	Disgust	Fear	Neutral
Happy	95.34	0	0	2.33	0	0	2.33
Sad	0	85.71	0	0	0	14.29	0
Surprise	2.44	2.44	87.8	0	0	4.88	2.44
Anger	0	0	3.45	89.65	0	3.45	3.45
Disgust	0	0	3.57	0	92.86	3.57	0
Fear	5.89	2.94	2.94	0	2.94	85.29	0
Neutral	0	0	6.9	0	0	6.9	86.2

4.4.2 Cohn-Kanade Database

The Cohn-Kanade database consists of images (frontal view) of 100 university students of different age, gender and ethnicity. The subjects were instructed to perform a series of 23 facial displays, six of which were based on description of basic emotions. The facial displays are converted to basic 6 expressions where possible and a total 348 image sequences are selected. Only the final image of each of the selected sequences are considered for our training and testing. The testing procedure is the same as mentioned in sub-section 4.4.2. The recognition rates given for this database in the experimental results are therefore the averages of the recognition rates of all tests conducted. The recognition rates for various combinations of curvelet (Scale, Orientation) and LBP (Neighbourhood, Radius) are discussed in Table 4.3. In this database also, curvelet (3,16) and LBP (8,2) combination gives the best results. But unlike the JAFFE database, for curvelet(2,8), LBP(8,2) gives the worst performance. From this disparity, it is clear that depending on the images, the trend exhibited

Table 4.3: Recognition rates obtained by different curvelet (scale, orientation) + LBP (neighbourhood, radius) combinations on Cohn-Kanade database

Combination Details	Recognition Rates (%)
curvelet(3,16) + LBP(8,3)	87.14
curvelet(3,16) + LBP(8,2)	90.33
curvelet(3,16) + LBP(8,1)	88.03
curvelet(3,8) + LBP(8,3)	87.14
curvelet(3,8) + LBP(8,2)	89.29
curvelet(3,8) + LBP(8,1)	88.55
curvelet(2,8) + LBP(8,3)	88.60
curvelet(2,8) + LBP(8,2)	86.81
curvelet(2,8) + LBP(8,1)	87.09

Table 4.4: Confusion matrix (%) obtained by curvelet (3, 16) and LBP (8, 2) on Cohn-Kanade database

Given Label \ Classified Label	Happy	Sad	Surprise	Anger	Disgust	Fear
Happy	83.87	3.23	4.35	1.02	0	7.53
Sad	3.33	88.33	1.67	1.67	3.33	1.67
Surprise	2.53	0	93.67	0	0	3.8
Anger	0	10.53	0	86.84	2.63	0
Disgust	0	2.38	0	0	95.24	2.38
Fear	6.97	2.33	4.65	0	0	86.05

by curvelet and LBP combinations may vary significantly. The confusion matrix obtained from this database for the said combination is presented in Table 4.4. From this matrix, it is clear that the confusion in Cohn-Kanade database is more than that of the JAFFE database. The most accurately detected expression is disgust. The least accurately detected expressions in this database are anger and fear.

4.4.3 Comparison with Other Methods

A comparison of the proposed method with Gabor wavelets based method and LBP based method [46] is shown in Table 4.5. The proposed feature extractor gives better performance compared to Gabor wavelets and LBP based method using the chi-square based nearest neighbour classifier. It is observed that the results improve more in the JAFFE database when compared to that in Cohn-Kanade Database. This may be attributed to the fact that Cohn-Kanade has more variations in ethnicity and gender compared to the JAFFE database. Thus, it is clear from the results that LBP on curvelet domain performs better than LBP in spatial domain.

Table 4.5: Comparison of proposed method with other methods

Methods	Classification Rates(%)	
	JAFFE	Cohn-Kanade
Gabor Wavelets	82.46	86.9
LBP based method	78.36	84.5
Proposed Method	93.69	90.33

4.5 Chapter Summary

A novel approach of using the curvelet transform and LBP for recognition of expressions from still face images is proposed. The experiments have been carried out on the popular JAFFE and Cohn-Kanade databases. Detailed experimental results have been shown for various combinations of curvelets and LBP using different scales, orientations, neighbourhood and radius. The method has showed promising results in JAFFE and Cohn-Kanade databases and this demonstrates the efficiency of curvelet based LBP as a better feature extractor than only LBP for facial expression recognition.

Chapter 5

Multiresolution Entropy for Facial Expression Recognition

5.1 Introduction

This chapter proposes the use of multiresolution entropy at selected points for classifying facial expressions from still images. A directional multiresolution transform like curvelet transform which refines its domain by using orientation information has been found suitable for expression classification in the earlier chapter. The method in the earlier chapter used only the approximate subband for feature extraction. But the high frequency information is contained in the detail subbands which have not been used before. Since similarity of facial expressions has been studied earlier using Gabor wavelet which uses filters oriented in different directions on specific feature points in images, the orientation selectivity and information content of curvelet subbands at specific facial points can prove to be better candidates as feature extractor and are used in this work. The information at selected facial points are gathered using the entropy of the corresponding pixel at various subbands. The method is evaluated in the JAFFE and Cohn-Kanade databases without and with cross-validations. Both of the cross-validated and non-cross-validated experimental procedures have been adopted in literature but very few works have addressed them simultaneously. The method is compared with wavelet entropy, contourlet entropy, Gabor wavelet and LBP based facial expression recognition methods. Experimental results show that the directional multiresolution subband entropy at selected points may be used to form effective

features for classifying facial expressions.

5.2 Framework for Recognition

The points selected on the face are the points that are most affected due to the six basic expressions of happiness, sadness, surprise, anger, disgust and fear. The coefficients at a specific point of the original signal in any subband of its MR are the measures of the strength of the point under the parameters of the corresponding subband. It has been already established [5, 48] that curvelets are very effective in representing edges and ridges. Hence, points around which edges and ridges occur due to expressions will have significant energy in specific subbands. Thus, the entropy at the selected point at any subband will differ due to different expressions. Earlier, wavelet energy and entropy obtained from all available subbands have been used for facial feature extraction [7, 8]. Since, the entire face has been considered for energy calculations in those methods, the global facial energy which has the details of a person's identity as well, is being used to generate the features. In order to eliminate that problem, the proposed approach is based on selecting 36 facial points from a face and analyzing the subband entropies at those points to form an effective feature set. Different neighbourhoods of a chosen point in curvelet domain are used to get the entropy. The extracted entropies are projected to PCA space. The vectors in the PCA space are subjected to Linear Discriminant Analysis (LDA). A simple nearest neighbour classifier is used for classification. As this work is based on facial features using dimension reduced curvelet entropy of selected points, the proposed method comprises of three main parts- facial feature points selection, calculation of subband entropy followed by dimension reduction and classification. All the images are subjected to tilt correction (the line connecting the centers of two eyes of a face is made horizontal by rotating the image) before selecting the facial points.

5.2.1 Facial Feature Point Selection

36 specific facial points are chosen manually from each of the faces for processing. 34 of those points are similar to those used in [61], 2 other points are used in this work. All the 36 points are shown in Fig. 5.1. Points marked in red are two additional points

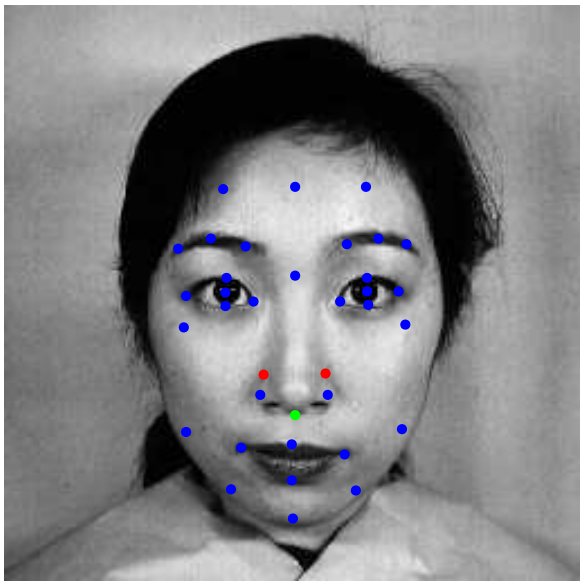


Figure 5.1: Selection of 36 facial points for feature extraction

considered for this work, as it is found by visual inspections that curves around these points vary significantly with different expressions. Also, the green point is a shifted form of the point at the tip of the nose owing to its changing proximity with the mouth for different expressions. The location of these points are stored for further processing.

5.2.2 Curvelet Subband Entropy

After the selection of 36 specific facial points, the face images are represented in curvelet domain with specific scale and orientation. Figure 5.2 shows the subbands of an image obtained by curvelet transform. A close look at the subbands reveal

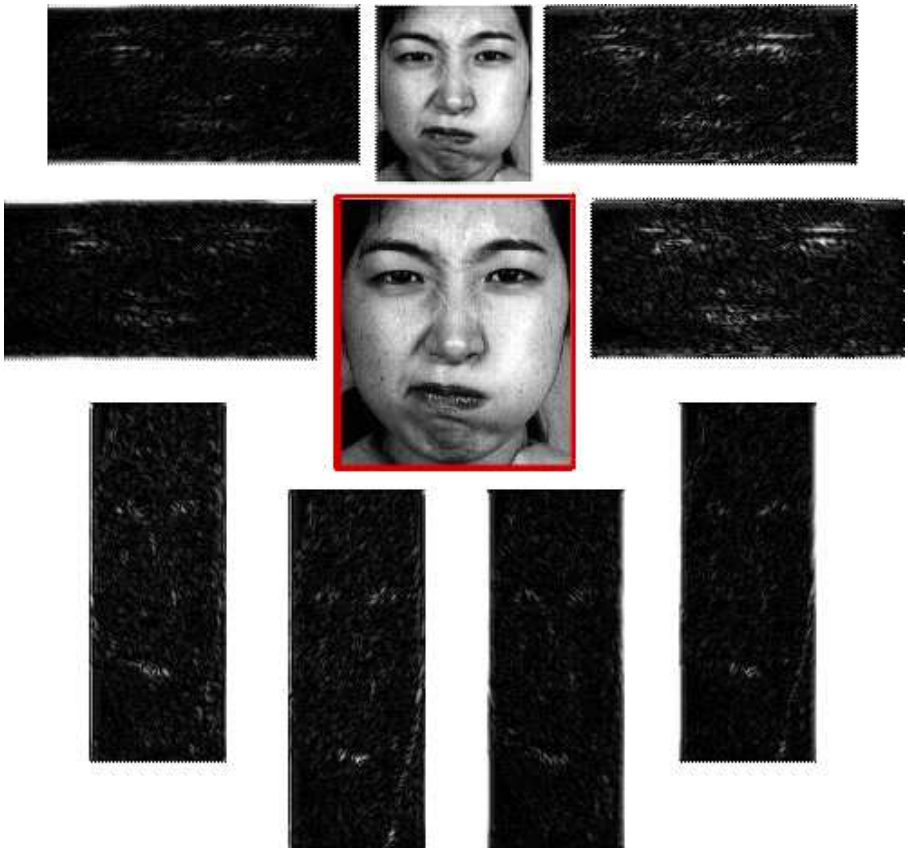


Figure 5.2: Curvelet subbands obtained from the image (bordered by red) at scale = 3 and orientation = 8

that the location and proportion of white patches (represents significant coefficients) varies in each subband. Therefore, it is evident from the figure that each subband has different information about the face. The number of subbands obtained depends on the scale (s) and orientation (o) of the curvelet transform. The location of the corresponding points of the selected/interest points are found out in each subband. All the curvelet subbands do not have same dimensions, hence the location of the corresponding point of an interest point varies in each subband. The corresponding points are identified in each subband. In each subband, centering each interest point, four energy values are calculated by taking square regions of sizes 3×3, 5×5, 7×7 and

9×9 according to eqn. 5.1.

$$E_W = \sum_W |S_n(x,y)|^2 \quad \text{for } W \ = \ 3,5,7 \text{ and } 9 \tag{5.1}$$

Here, $S_n(x,y)$ is the coefficient at the point (x,y) in subband S_n. The energy values E_W are normalized by dividing them by the total energy E_T of the original image I where

$$E_T = \sum_I |I(p,q)|^2 \tag{5.2}$$

The normalized energies are denoted by E_{WN} where

$$E_{WN} = \frac{E_W}{E_T} \quad \forall\ W \tag{5.3}$$

The normalized energies thus become higher as the W increase. This is obvious for all the interest points. In order to have an idea about the neighbouring coefficients with the increase of W, normalized energy per point E_{WNP} is calculated for all window sizes by eqn. 5.4.

$$E_{WNP} = \frac{E_{WN}}{W^2} \quad \forall\ W \tag{5.4}$$

Here, the distribution of the energy of an interest point is obtained by E_{WNP}. The value will be higher if the corresponding point in any subband is surrounded by more number of significant coefficients. The peakiness of this energy distribution is formulated by entropy which highlights uniform energy variation in all windows when large in value. Same expressions will yield similar orientation curves which implies similar variation in energy distribution surrounding the interest points in same subband. Therefore, entropies of interest points are used here. Before calculating the entropy, the sum of all energy values obtained for a point is made 1 by eqn. 5.5

$$E_{WNPN} = \frac{E_{WNP}}{\sum_{W=3,5,7,9}(E_{WNP})} \tag{5.5}$$

Finally, the entropy for any interest point is calculated by the following equation

$$ENT_{i,j} = \sum_W E_{WNPN} \log_2 E_{WNPN} \tag{5.6}$$

Thus, from each subband 36 values of entropy will be obtained. If N subbands are present, the feature vector for an image is obtained by concatenating the entropy values from all subbands. Hence, the size of the feature vector is $36 \times N$. During comparison with contourlets and wavelets, the contourlet and wavelet entropies are calculated instead of the curvelet entropy and then subjected to dimension reduction methods.

5.2.3 Dimension Reduction and Classification

The aim of dimension reduction methods is to reduce the number of variables (or dimensions) required to represent the data while retaining the variations present in the data set. The feature vectors obtained using multiresolution entropy are large in size and hence dimension reduction methods is necessary. The expression recognition task performed here classifies images into one of the 7 (6 basic expressions + 1 neutral) classes. One of the most popular dimension reduction method is PCA but it suffers from a basic problem. It does not care about the separability of classes. Let X be a training feature set of size $n \times d$, where n is the number of training vectors and d is the dimensionality of each vector. When X is subjected to PCA, it directly works with the zero-mean version Y of X to generate the scatter or covariance matrix S_c of size $d \times d$. The eigen vectors corresponding to the eigen values of S_c are the principal components and they are uncorrelated. Generally, the principal components are ordered in a matrix P_{mat} in columns, so that they account for the decreasing amount of variation in the data. P_{mat} has size of $d \times r$, where r is the number of principal components selected. The principal components are basis vectors (or axes) such that their linear combination can provide a description of the original data in the PCA subspace. The product $P_{val} = Y.P_{mat}$ of size $n \times p$ is called the PCA projection of the training data. Evidently,P_{val} is generated without taking care of the classes to which the training samples belong to. Therefore though the projection axes chosen by PCA account for data variation, they may not be able to provide good discrimination.

The solution to the problem of PCA was given by LDA which works by discriminating between the classes. LDA works to find directions in which the classes are separated maximally. For generality, let the matrix X of size $n \times d$ has n feature vectors in its rows which can be classified to C classes. Let, μ_i be the mean vector of length d (equal to the number of columns in X) of i^{th} class where $i = 1, 2, 3, ..., C$. The mean of all classes is $\mu = \frac{1}{C}\sum_{i=1}^{C}\mu_i$. Let n_i be the number of samples within class i, where $i = 1, 2, 3, ..., C$. Therefore, the total number of samples is equal to the number of rows n of X, such that $n = \sum_{i=1}^{C} n_i$. Two covariance or scatter matrices are calculated which are called $within - class\ scatter\ matrix\ S_w$ and $between - class\ scatter\ matrix\ S_b$. S_w and S_b which are calculated as follows:

$$S_w = \sum_{i=1}^{C}\sum_{j=1}^{n_i}(x_{i,j} - \mu_i)(x_{i,j} - \mu_i)^T \tag{5.7}$$

$$S_b = \sum_{i=1}^{C} n_i(\mu_i - \mu)(\mu_i - \mu)^T \tag{5.8}$$

LDA attempts to find a linear transformation to maximize $\frac{det(S_b)}{det(S_w)}$. It is found that the linear transformation is given by L_{mat} whose columns are eigenvectors of $S_b{S_w}^{-1}$. At most, C-1 non-zero generalized eigenvectors could be found for the C-class LDA. The problem arises in the maximization, when the S_w is singular. As the number of training samples are lesser than the length of the multiresolution entropy vector obtained from each image, a singular S_w will be generated. Thus, the vectors are projected to PCA space with number of principal components lesser than the number of training images at first. The data P_{val} in PCA space is projected to LDA space by

$$L_{val} = L_{mat}^T P_{val} - mean(P_{val}). \tag{5.9}$$

The L_{val} is used for training. When a query image is obtained, multiresolution entropy is calculated from it and projected to PCA + LDA space. A nearest neighbour classifier with Euclidean distance as the metric is used to classify the query image. The process diagram of the proposed method can be summarized as shown in Fig. 5.3.

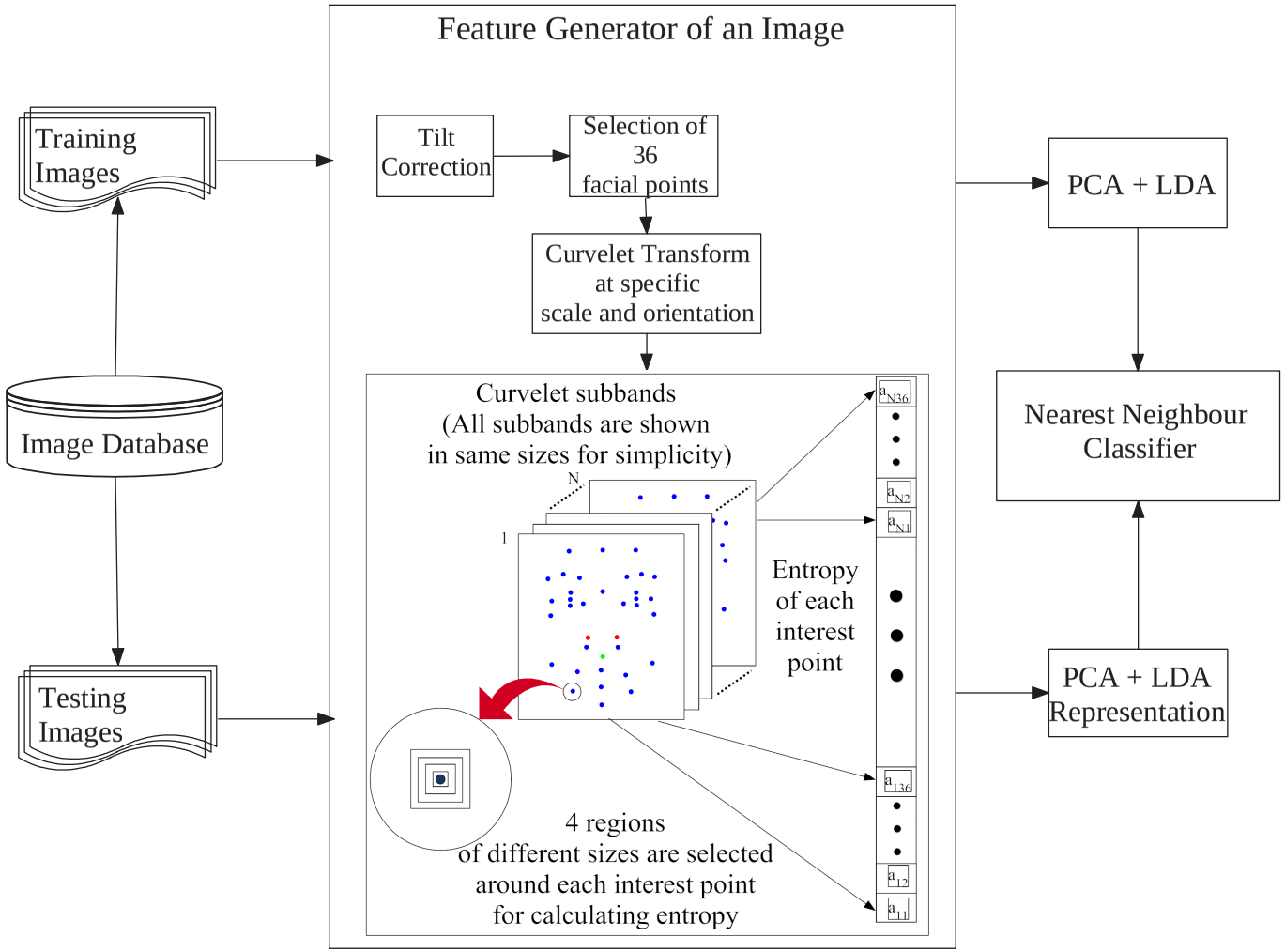


Figure 5.3: Curvelet entropy based facial expression recognition

5.3 Experiments without Cross-validation

The experiments without cross-validation here implies that when the images of a database are divided into training and test sets, there exist common subjects between the sets. In the JAFFE and Cohn-Kanade databases the experiments are carried out with curvelets, contourlets and wavelets.

5.3.1 Results in JAFFE Database using Curvelets, Contourlets and Wavelets

For JAFFE database, each subject has at least one image of each expression. Therefore, by taking 7 different expressions (including neutral) from each subject a test set of 70 images is formed. Rest of the images are used for training. This testing scheme is the same as used in [1]. This procedure of generating the test image set is done five times and it is ensured that every image of the database becomes a part of the testing set for at least once. Fig. 5.4 shows the recognition results in JAFFE database. For curvelets and contourlets, different scales and orientations are used. For wavelets, 3 levels of decomposition are used. Keeping the multiresolution parameter constant, the recognition rates are plotted against the number of principal components. To keep the S_b non-singular, a maximum of 130 principal components(less than the number of images in the training set) are considered. 5 different training sets are shown in different colours. From Fig. 5.4, the accuracy of recognition increases with the increase of number of orientations while keeping the scale constant for curvelets. For results with contourlets also, as shown in Fig. 5.5 the accuracy improves with the number of orientations. As shown in Fig. 5.6, for wavelets, the recognition rate gets better with the increase of level of decomposition. The rates obtained by curvelets are better than contourlets and wavelets.

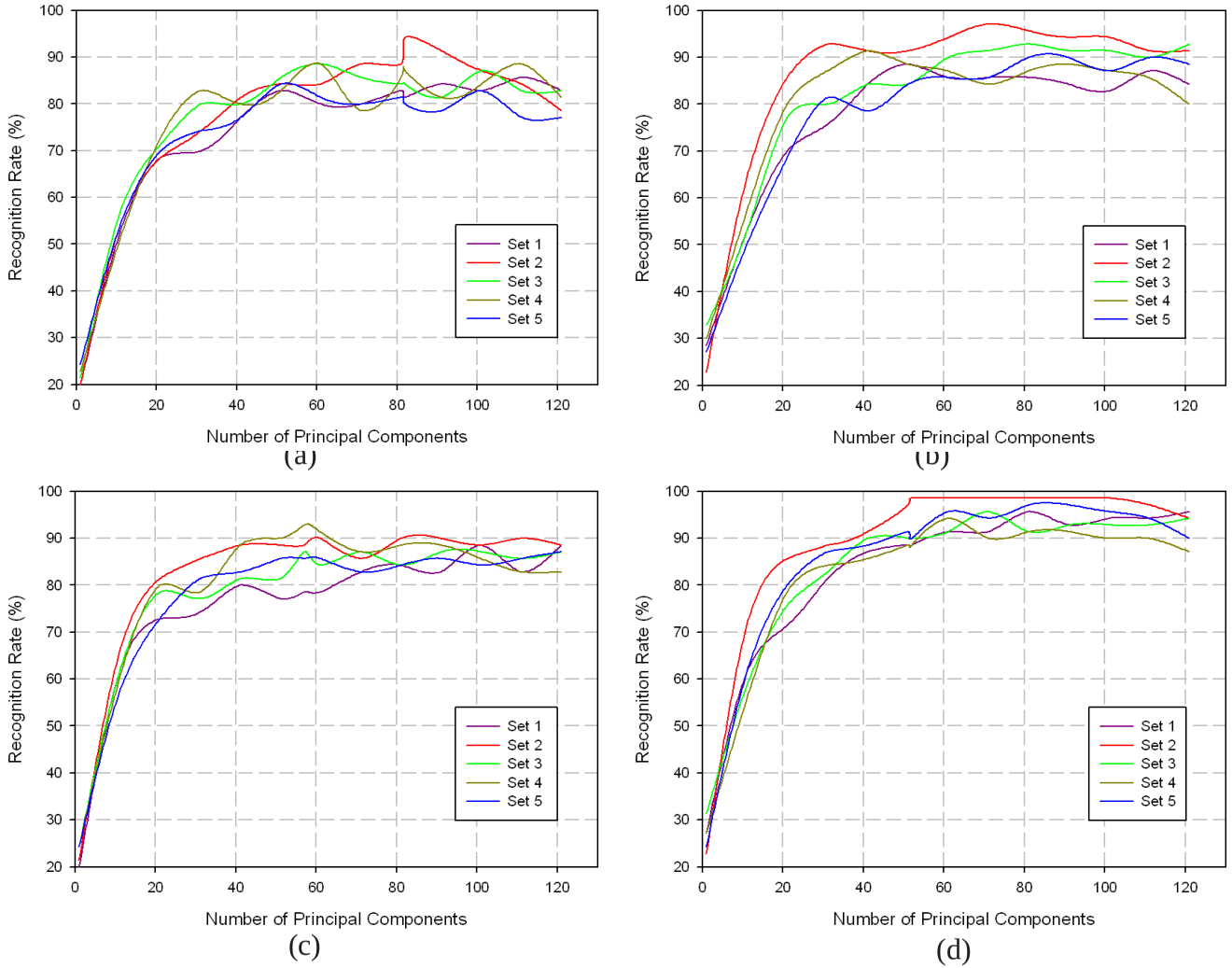


Figure 5.4: Recognition rates in JAFFE database by curvelets for (a) $s = 3$, $o = 16$, (b) $s = 3$, $o = 32$, (c) $s = 4$, $o = 8$, (d) $s = 4$, $o = 16$

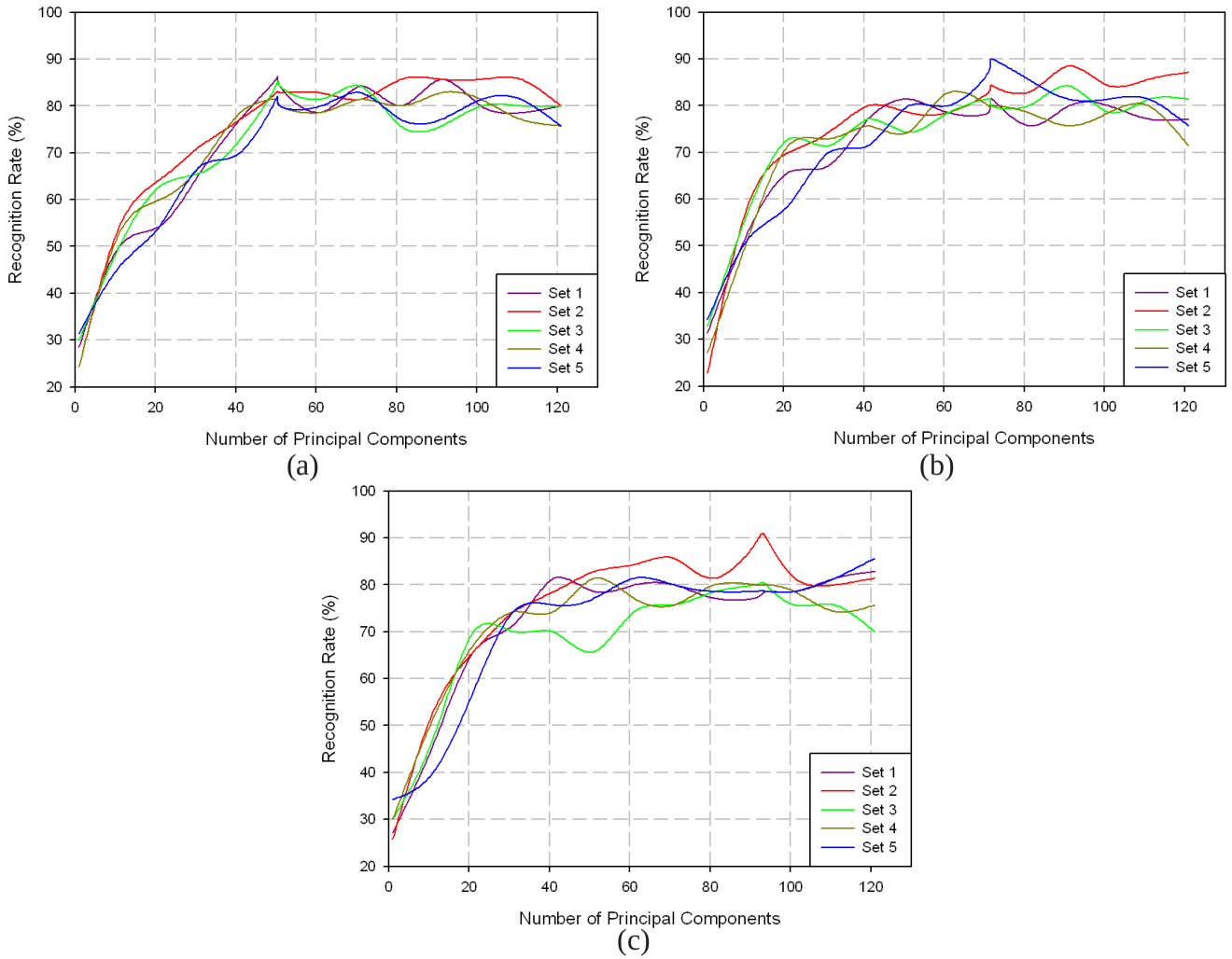


Figure 5.5: Recognition rates in JAFFE database by contourlets for (a) $s = 4$, $o = 16$, (b) $s = 4$, $o = 22$, (c) $s = 4$, $o = 28$

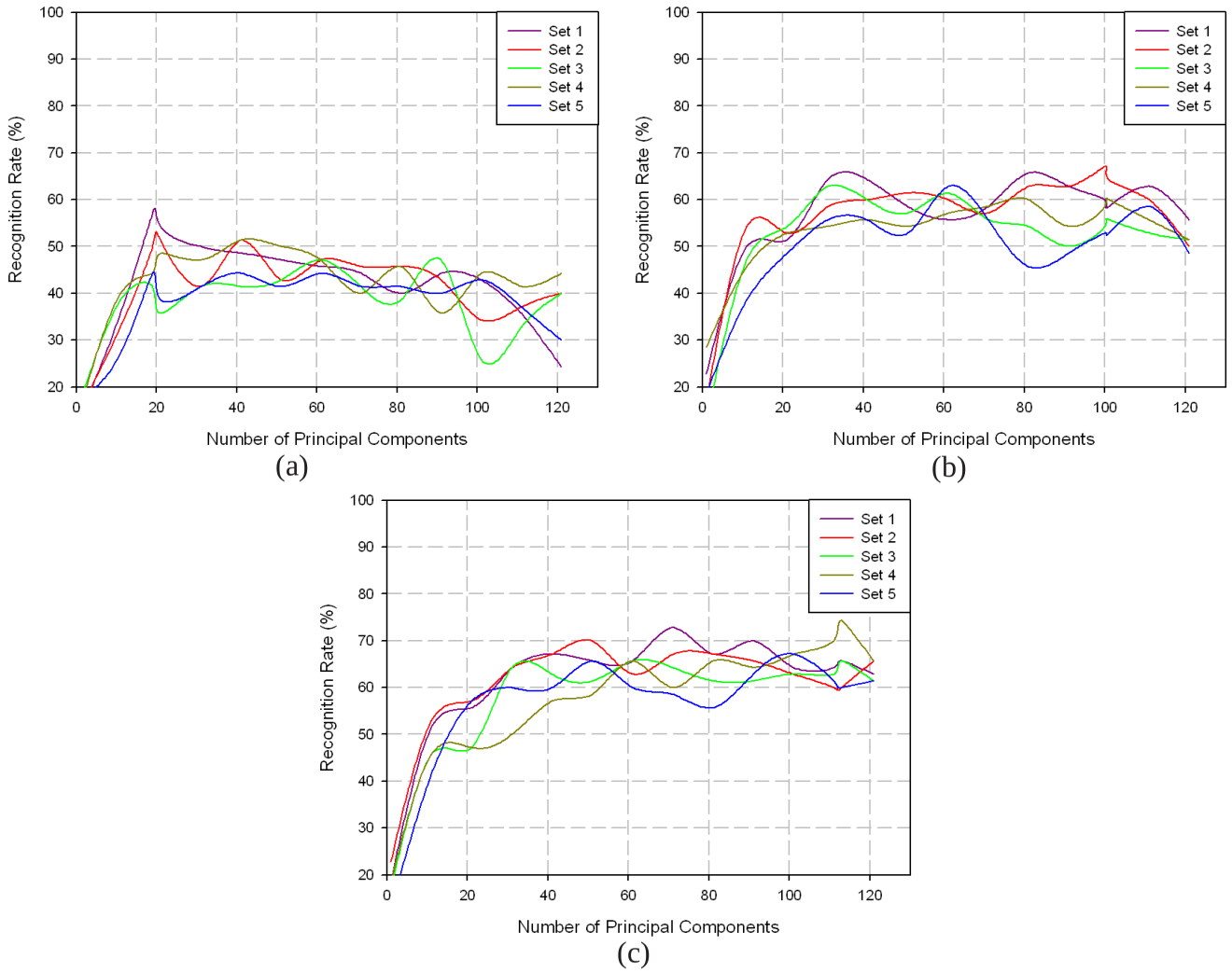


Figure 5.6: Recognition rates in JAFFE database by wavelets for (a) level 1, (b) level 2, (c) level 3

5.3.2 Results in Cohn-Kanade Database using Curvelets, Contourlets and Wavelets

86 subjects of Cohn-Kanade database are selected such that none of the facial points are occluded and total 355 images are obtained. The testing set of Cohn-Kanade database is formed by taking one image of every subject such that each test set has roughly same number of images of each of 7 expressions. Rest of the images are used for training. Total five different test sets are formed to ensure that each image of the database falls in the testing set for at least once. The test set sizes lie between 70-72. Though the number of training images are more than that of the JAFFE database, the maximum number of principal components is kept 130 for an uniform comparison of results between the two databases.

It is observed from the Figs. 5.7, 5.8 and 5.9 that the recognition accuracy is remarkably lower than that of JAFFE database. It may be attributed to the fact that since, instances of all expressions of a subject are not present in the training set (unlike JAFFE), the training set needs to be increased for better results. Also, the recognition rates obtained using curvelets, contourlets and wavelets are similar. Use of directional multiresolution transforms could not provide much better results than wavelets. But, the maximum accuracy is given by contourlets here. The first test set gives the poorest performance among all sets. This trend of the first set is exhibited by all of the three multiresolution methods.

5.4 Experiments with Cross-validation

The experiments with cross-validation implies that an intentional mismatch has been kept between the test set and train set. The mismatch may be in the subjects considered, the resolution of the images or their preprocessing. Here, cross-validation means that a subject independent facial expression recognition is targeted. The training sets

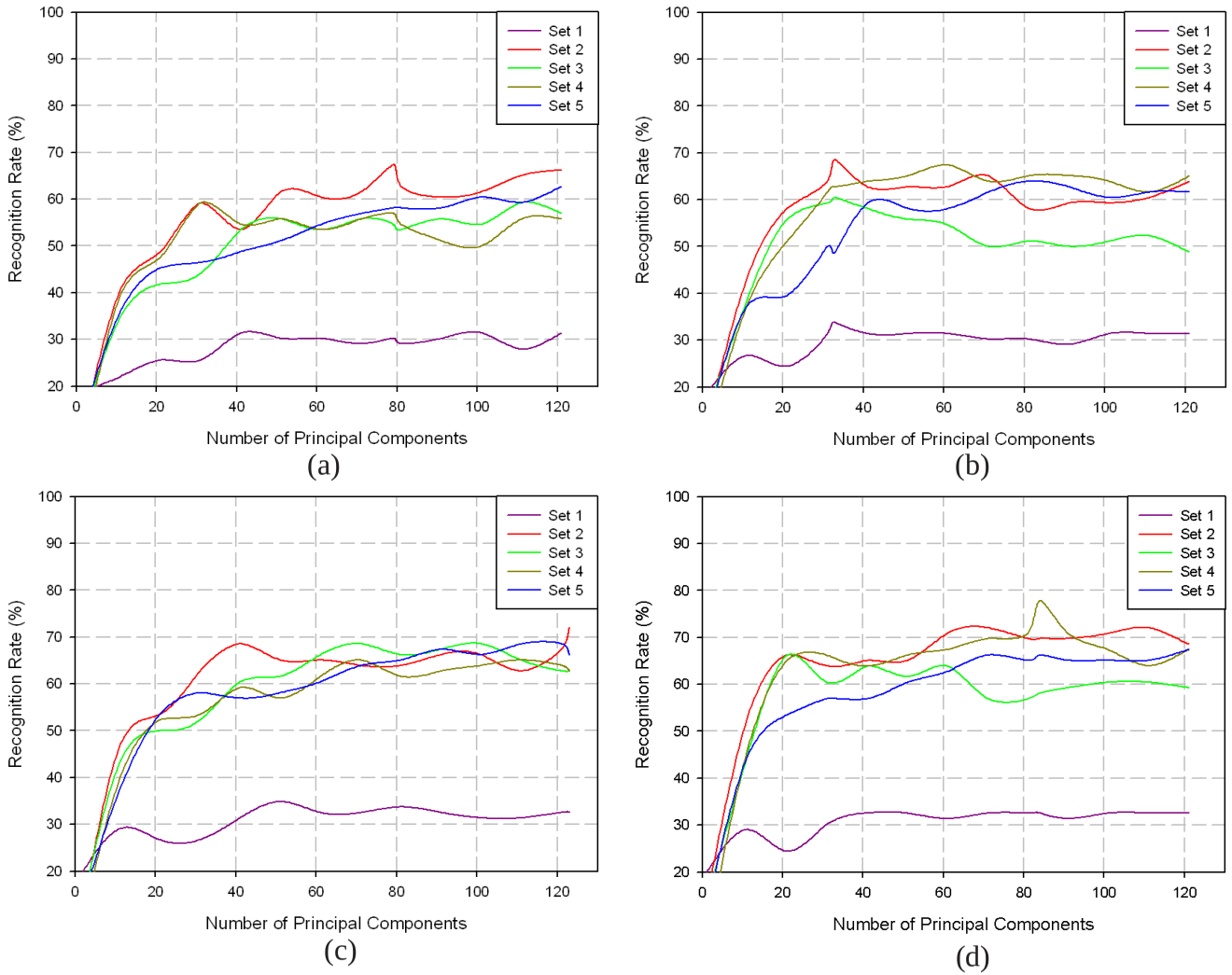
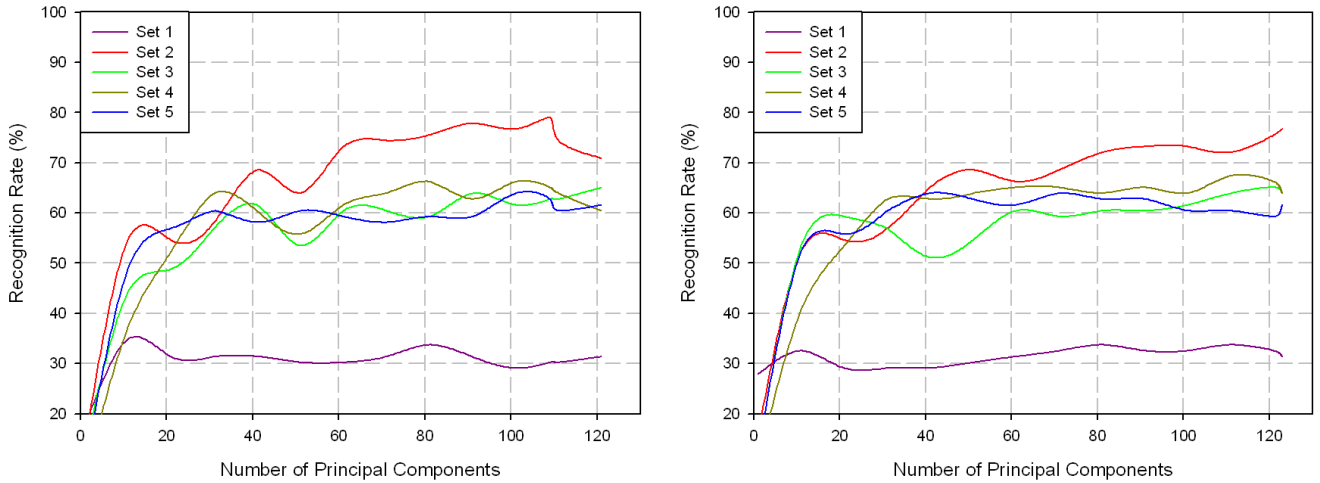


Figure 5.7: Recognition rates in Cohn-Kanade database by curvelets for (a) $s = 3$, $o = 16$, (b) $s = 3$, $o = 32$, (c) $s = 4$, $o = 8$, (d) $s = 4$, $o = 16$



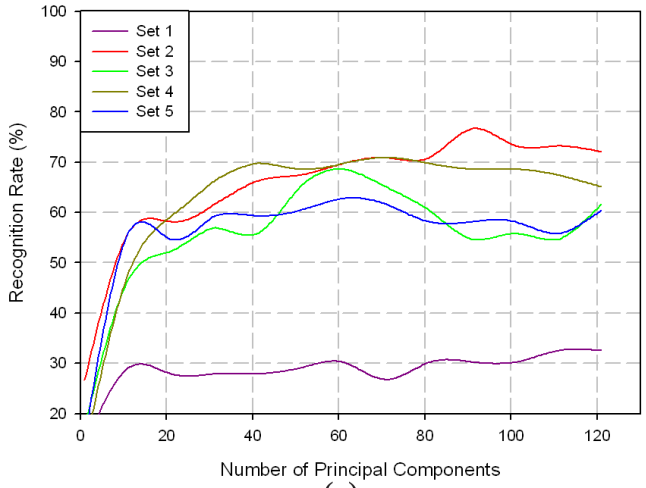


Figure 5.8: Recognition rates in Cohn-Kanade database by contourlets for (a) $s = 4$, $o = 16$, (b) $s = 4$, $o = 22$, (c) $s = 4$, $o = 28$

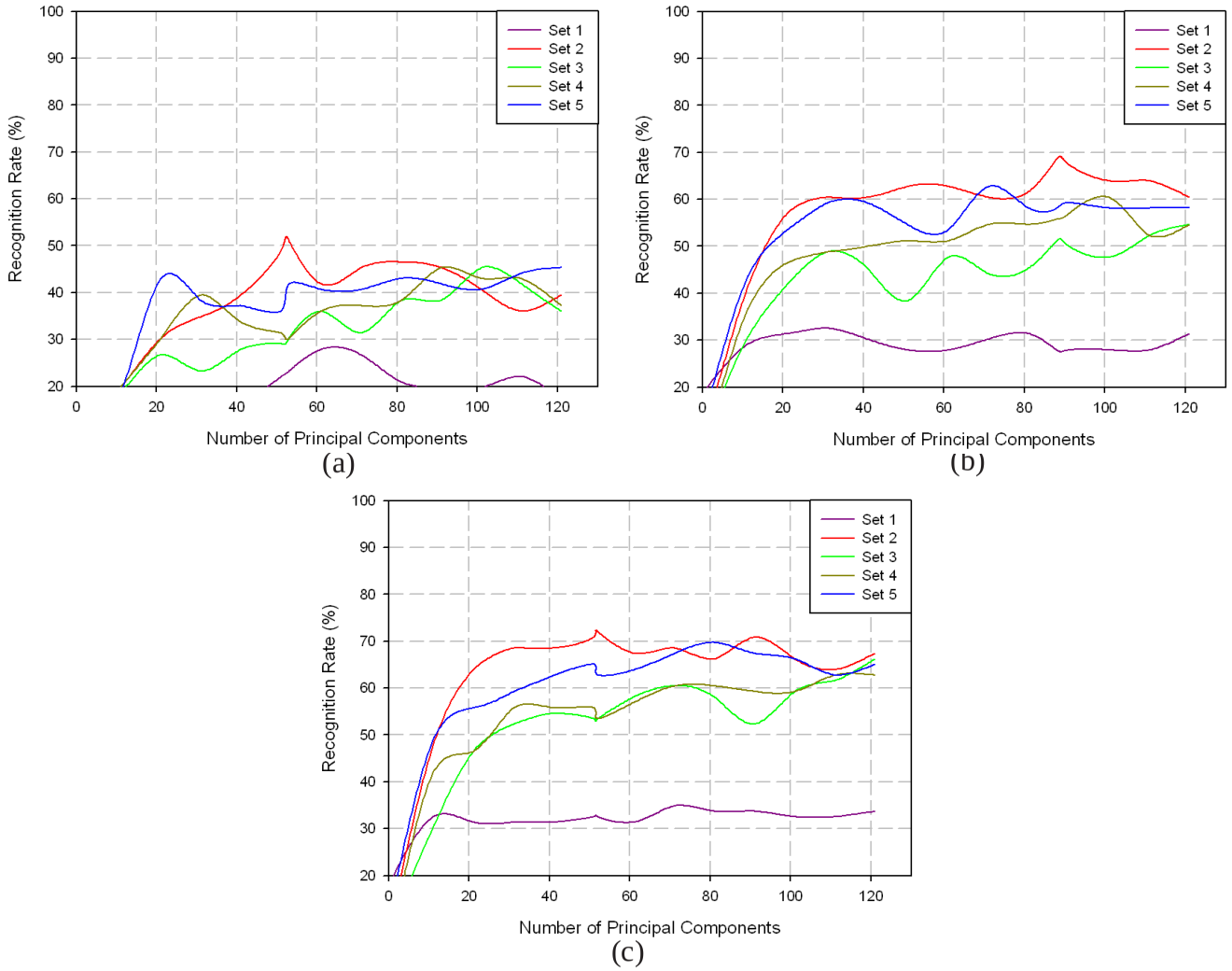


Figure 5.9: Recognition rates in Cohn-Kanade database by wavelets for (a) level 1, (b) level 2, (c) level 3

and test sets in which the databases are divided share no common subject. Experiments with cross-validation are important as they point out how well an expression is modeled being independent of a particular face. This testing is carried out in literature to test the generalized performance of any facial expression recognition method.

5.4.1 Results in JAFFE Database using Curvelets, Contourlets and Wavelets

For the JAFFE database, 10-fold cross-validations are used and it is similar to *leave one out* strategy since only 10 subjects are present. In each round, all images of a subject are used for testing and images of 9 other subjects are used for training. The test set size lie between 20-23. The recognition rates are shown in Fig. 5.10 for curvelets and it is clear that the results depend greatly on the subject under testing. The recognition rates are quite irregular in this database. Thus, the average recognition rate is pretty low which is also shown in [46]. As exhibited in Figs. 5.11 and 5.12, the tendencies of recognition rates among different test sets are very dissimilar. Thus, recognition rates obtained by using curvelets, contourlets and wavelets are found to exhibit the same fluctuating nature. These results conform with the non-uniformness of expressions present in the JAFFE database as pointed out by some earlier works in literature.

5.4.2 Results in Cohn-Kanade Database using Curvelets, Contourlets and Wavelets

A 10-fold cross-validation is conducted in Cohn-Kanade database as well. In each round, the testing set consists of 8-9 subjects and number of test images lie between 33-49. The results are shown in Fig. 5.13. The results obtained are better than that of the non-cross-validated ones. This supports that with enough training images, the proposed method is going to perform better. Also, recognition rates for different sets

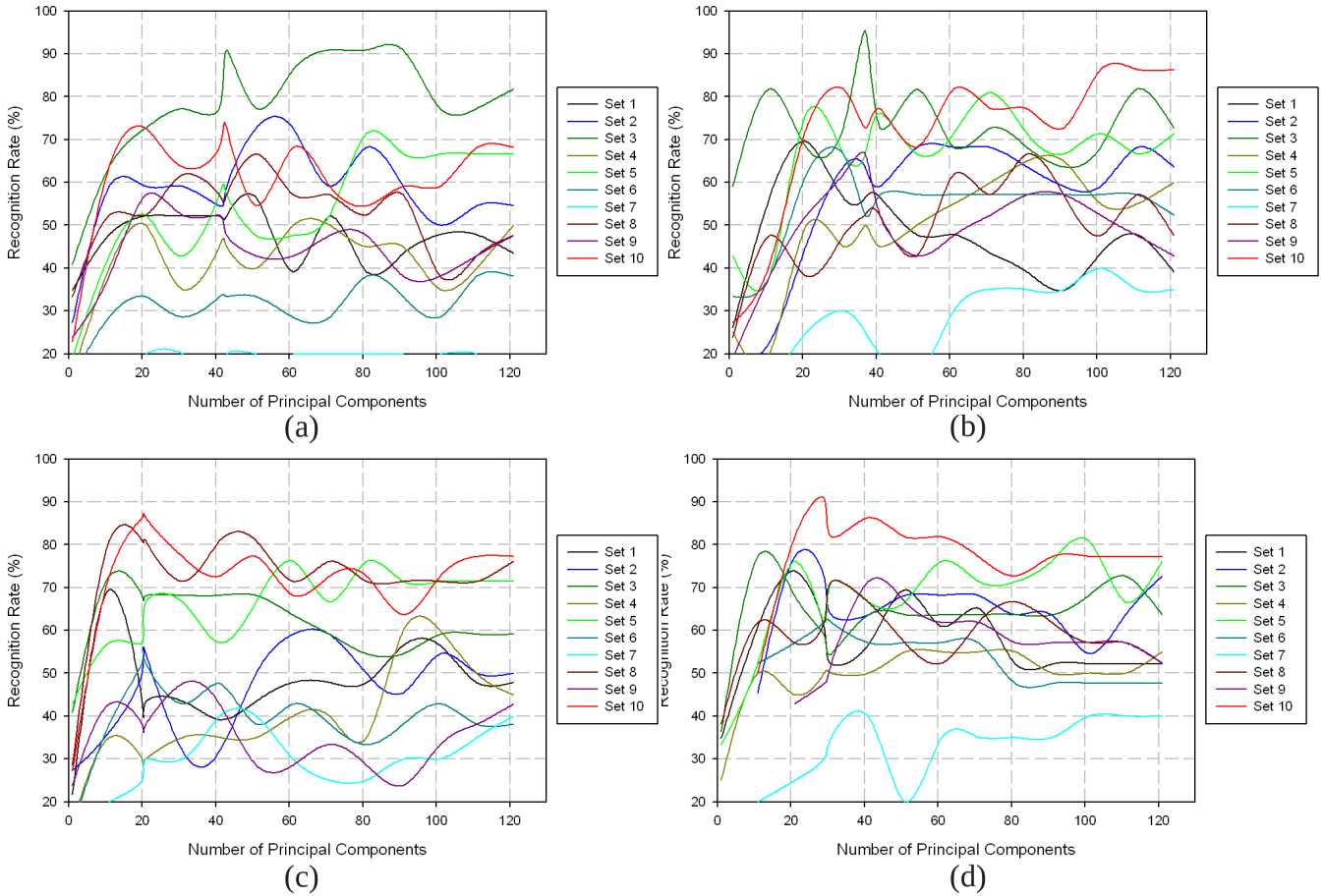


Figure 5.10: Cross-validated recognition rates in JAFFE database by curvelets for (a) $s = 3$, $o = 16$, (b) $s = 3$, $o = 32$, (c) $s = 4$, $o = 8$, (d) $s = 4$, $o = 16$

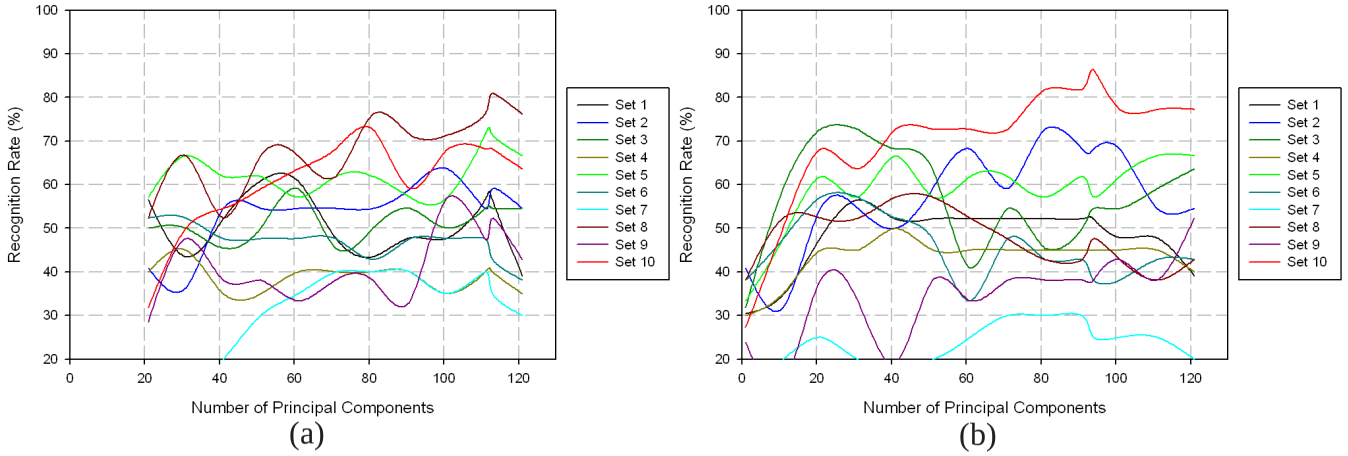


Figure 5.11: Cross-validated recognition rates in JAFFE database by contourlets for (a)$s = 4$, $o = 16$, (b)$s = 4$, $o = 22$, (c)$s = 4$, $o = 28$

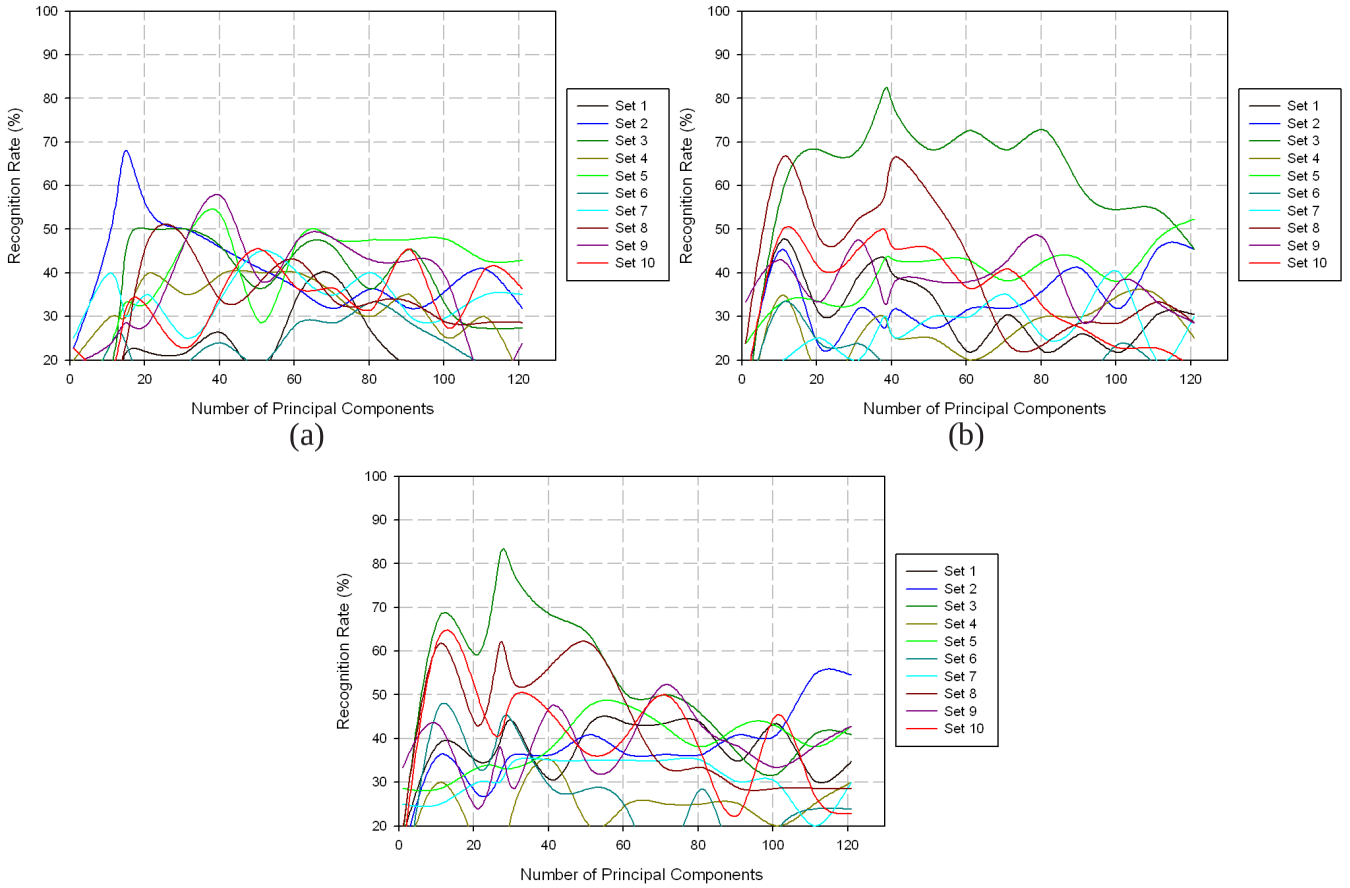


Figure 5.12: Cross-validated recognition rates in JAFFE database by wavelets for (a) level 1, (b) level 2, (c) level 3

are quite regular compared to that in the JAFFE database. The cross-validated results indicate how well the method can perform to recognize face independent expressions. This implies that regularity or uniformity in the expressions is also very important. Therefore, more number of training images can improve results significantly. As already expressed in [46]. The small size of the JAFFE database may be responsible for low and less regular recognition rates. More training images in Cohn-Kanade database give better cross-validated results.

5.5 Comparison with Gabor Wavelets and LBP

In this part the curvelet based entropy is compared with gabor wavelet based entropy and LBP + PCA + LDA [46] method which is one of the best methods in literature. For gabor wavelets, 6 orientations of gabor filter with wave numbers $\pi/4$, $\pi/8$ and $\pi/16$ are used and this is similar to [61]. For LBP, uniform LBP(8,2) operator is used. Both cross-validated and non-cross-validated experiments are carried out in JAFFE and Cohn-Kanade database using the same sets of train and test images as mentioned in sub-sections 5.3 and 5.4. The results are presented in Figs. 5.16 and 5.17. It is observed that with gabor wavelet entropy and LBP, the trend is similar to what is obtained with curvelets.

It is found from 5.16, that directional multiresolution entropy performs better than Gabor wavelet entropy, though the general trend of the recognition rates are similar. For non-cross-validated experiments in JAFFE database, the recognition rates are higher and uniform. The cross-validated rates in JAFFE are irregular and much lower with Gabor wavelets. For Cohn-Kanade database the recognition rates for cross-validated and non-cross-validated experiments are irregular though cross-validated recognition rates are comparatively more uniform. Similar trend is found using LBP, though the rates are much better than that obtained using Gabor wavelets. Curvelet entropy is comparable to LBP based recognition. For non-cross-validated

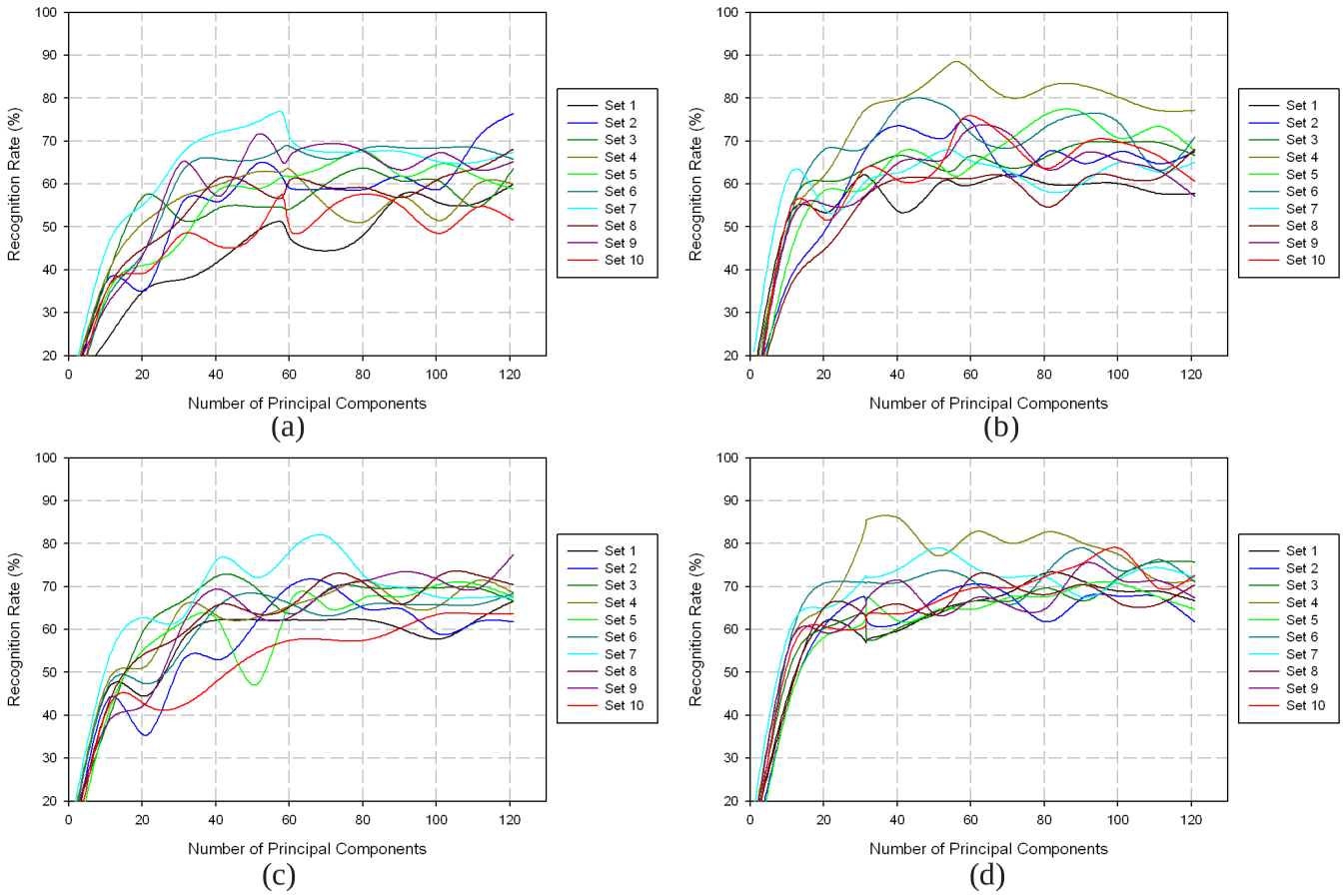


Figure 5.13: Cross-validated recognition rates in Cohn-Kanade database by curvelets for (a) $s = 3$, $o = 16$, (b) $s = 3$, $o = 32$, (c) $s = 4$, $o = 8$, (d) $s = 4$, $o = 16$

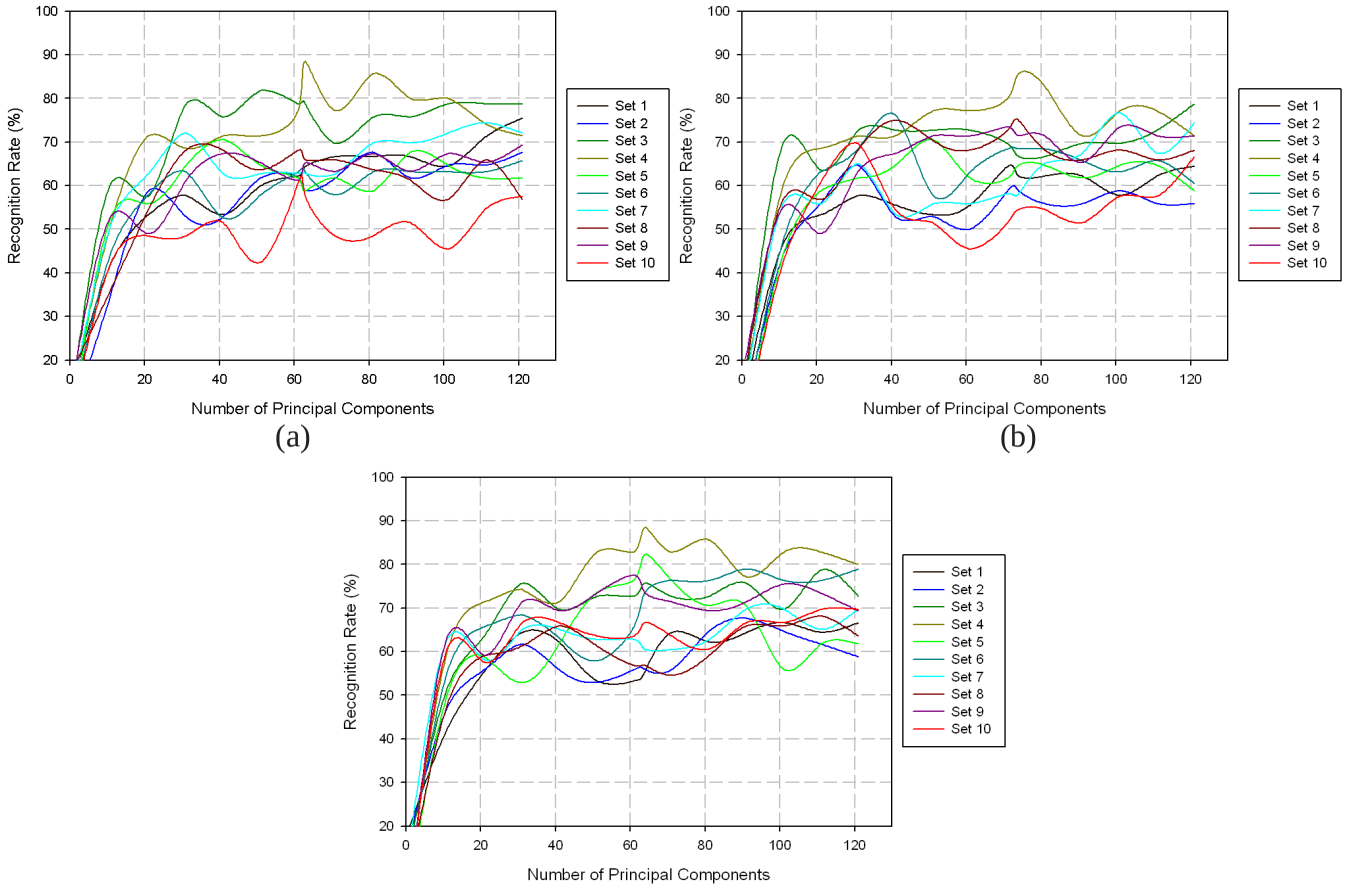


Figure 5.14: Cross-validated recognition rates in Cohn-Kanade database by contourlets for (a)$s = 4$, $o = 16$, (b)$s = 4$, $o = 22$, (c) $s = 4$, $o = 28$

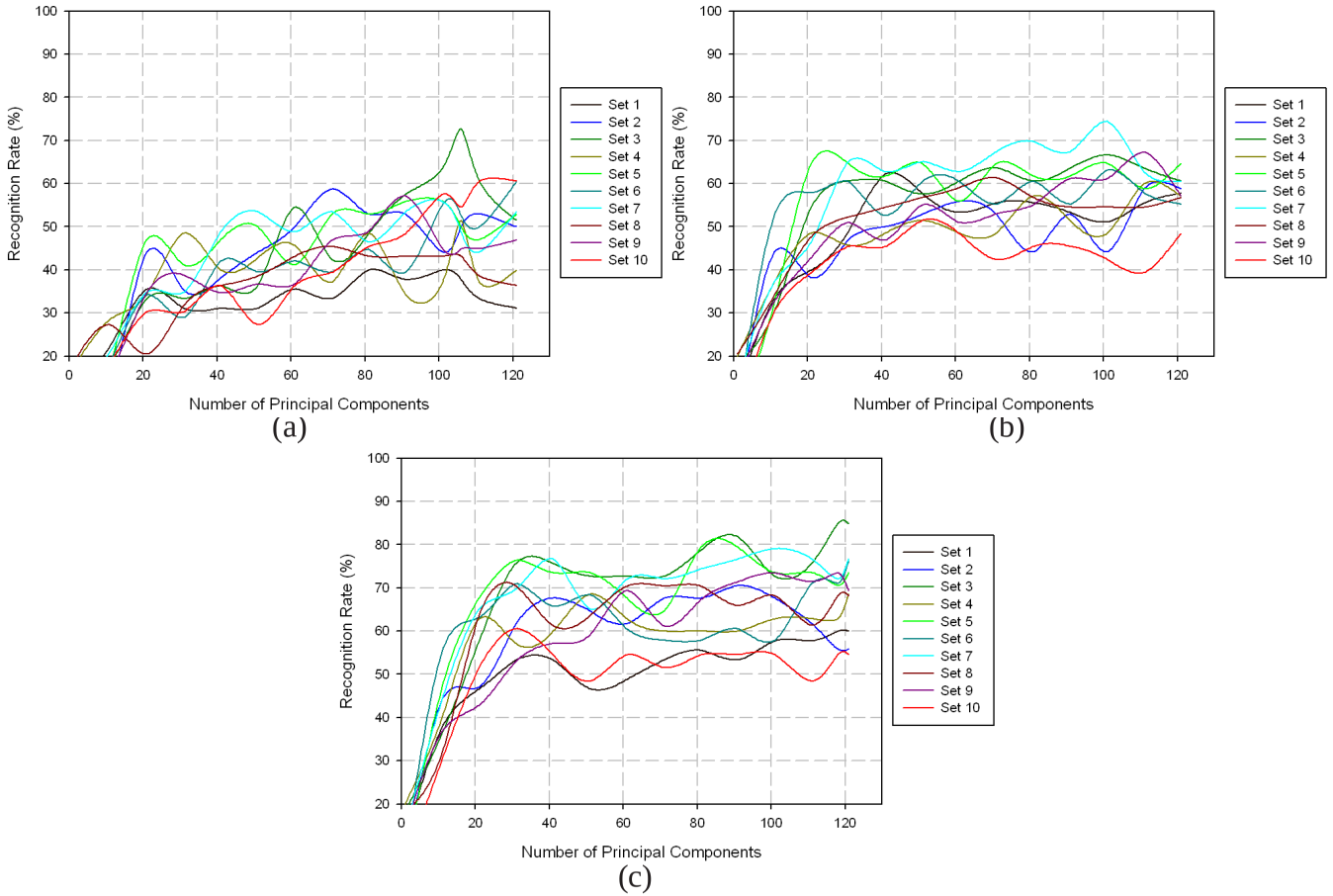


Figure 5.15: Cross-validated recognition rates in Cohn-Kanade database by wavelets for (a) level 1, (b) level 2, (c) level 3

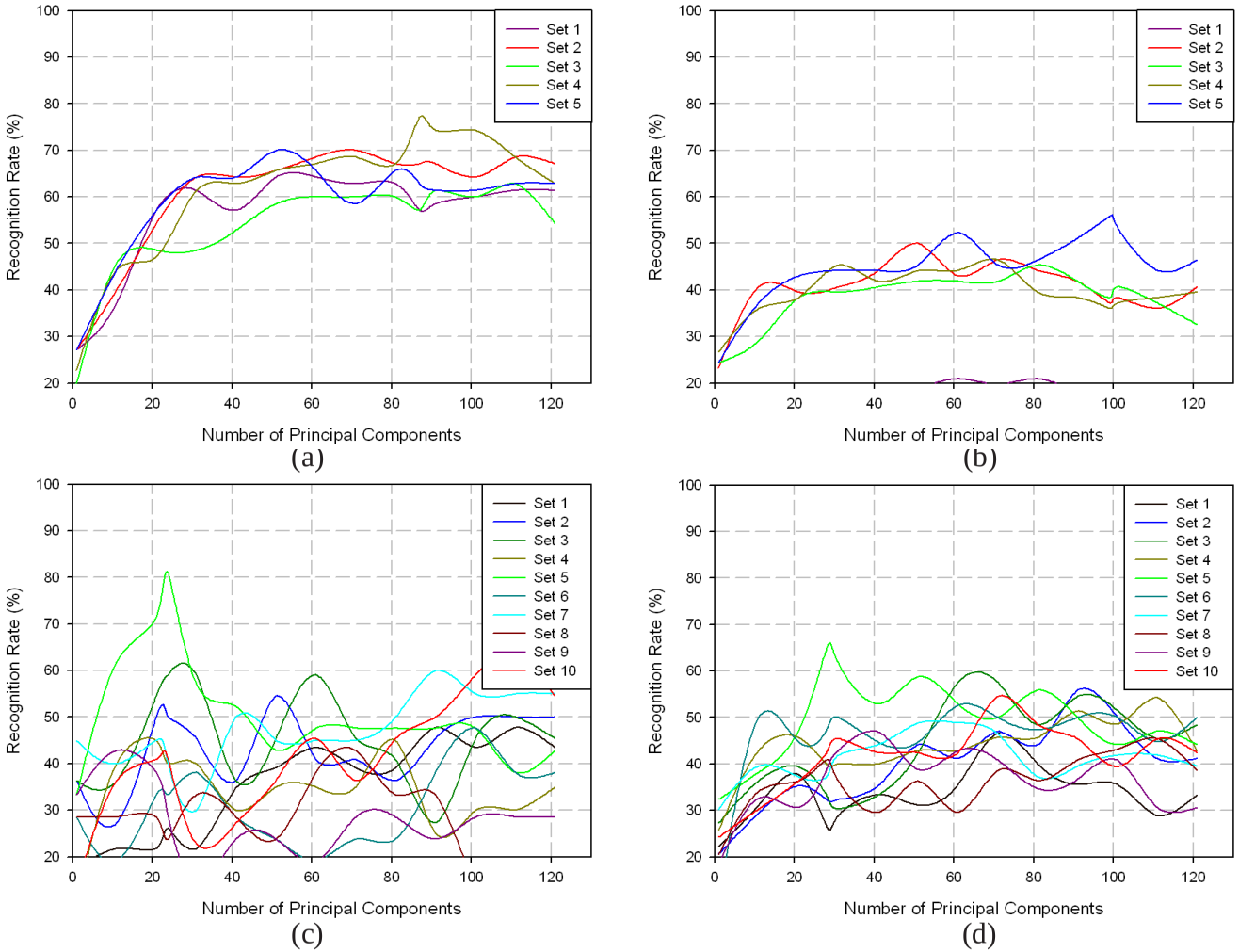


Figure 5.16: Gabor entropy based recognition in (a) JAFFE non-cross-validated, (b) Cohn-Kanade non-cross-validated (c) JAFFE cross-validated and (d) Cohn-Kanade cross-validated

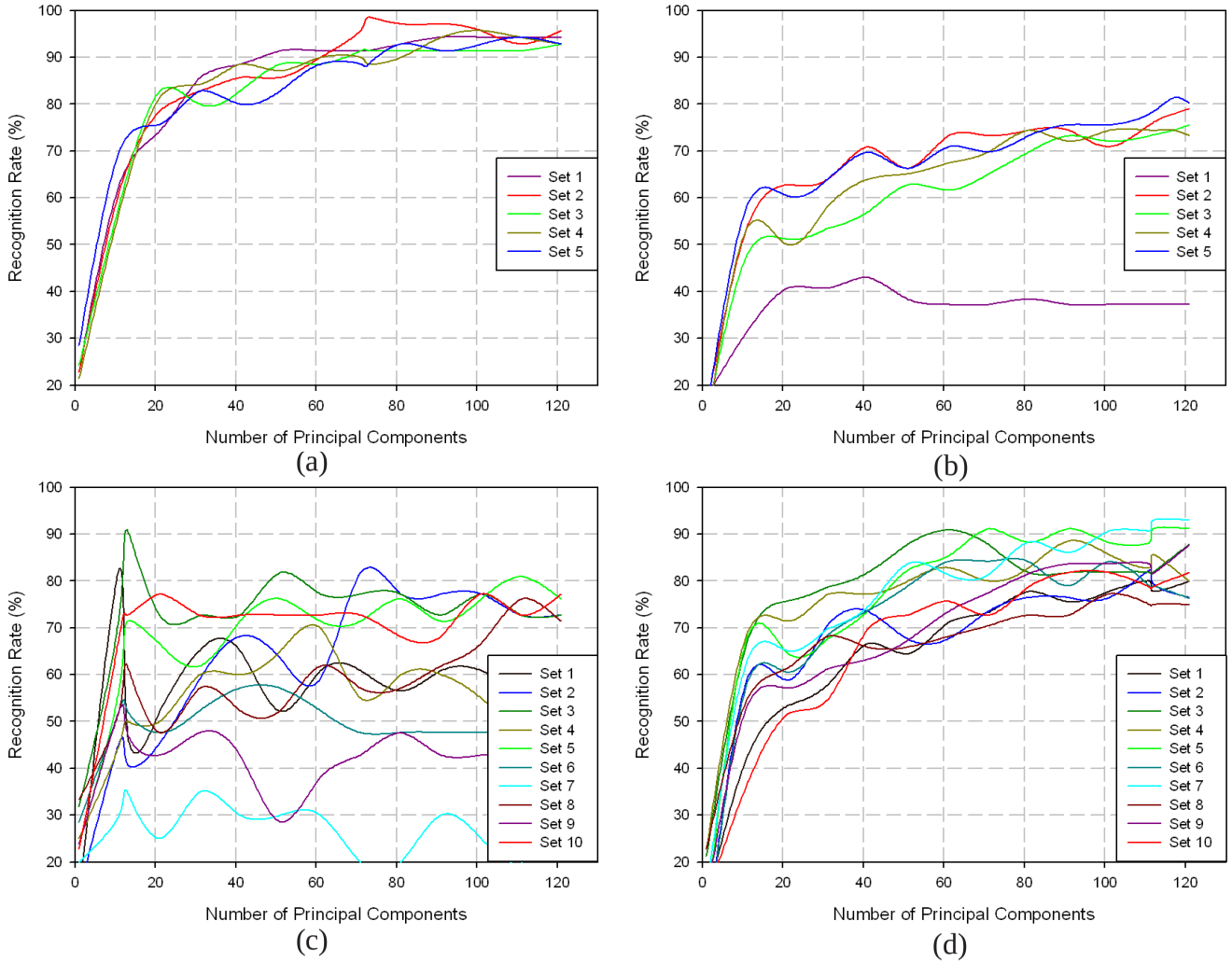


Figure 5.17: LBP + PCA + LDA based recognition in (a) JAFFE non-cross-validated, (b) Cohn-Kanade non-cross-validated (c) JAFFE cross-validated and (d) Cohn-Kanade cross-validated

experiments in JAFFE database, directional multiresolution performs better than LBP whereas for cross-validated experiments in Cohn-Kanade, LBP performs better. The maximum recognition rates offered by curvelet entropy for non-cross-validated experiments in Cohn-Kanade and cross-validated experiments in JAFFE is higher than what is achieved by LBP + PCA + LDA.

5.6 Chapter Summary

A method for facial expression recognition using multiresolution entropy has been proposed. The main idea is to use curvelet subbands at selected facial points and use the curvelet entropy as features for expression. The proposed method has also been verified using contourlet and wavelet entropy. Extensive results have been given on non-cross-validated and cross-validated experiments conducted on the JAFFE and Cohn-Kanade databases. The robustness of the method against age, gender and illumination variation is established since, each of the databases has some of these variations. The results show that the proposed method is better than Gabor wavelets and is comparable with LBP based method for facial expression recognition.

Chapter 6

Drowsiness Detection using Multiresolution Entropy

6.1 Introduction

Drowsiness detection is the process of identifying a sleepy person. The automatic detection of drowsiness has gained interest in research community due to its practical importance. As pointed out by Traffic Injury Research Foundation of Canada, 4% of fatal crashes and 20% of non-fatal crashes take place due to the drowsiness of the driver. Thus, substantial amount of research is going on to detect drowsiness by using machines, to avoid such disasters. There are mainly two kinds of approaches that are followed: (a) Physiological signal analysis (b) Visual Data Analysis. In physiological signal analysis, mainly, the EEG (Electroencephalogram) spectra of a person or subject is utilized to make a decision about the state/level of drowsiness [25, 57] of the person. The approach followed in visual data analysis uses image, image sequences or videos for drowsiness detection. In this chapter, a study is made to analyze how well the facial features (discussed in chapter 5) are suited to differentiate between drowsy and non-drowsy faces.

6.2 Background

The visual data based algorithms focus mainly on the eyes of the subject. Most approaches rely on the method of Percentage of Eye Closure (PERCLOS). PERCLOS is a method for eye blinking measurement and it is the percentage of time during which

the eye is closed in a given time period. A sleepy person blinks less number of times in a given time limit but takes more time to complete a blink [58]. It is found that high PERCLOS scores are correlated to drowsiness. It has been the aim of many research works to measure eye blinks to obtain PERCLOS scores for classification [51, 26, 52]. Other visual data based approaches use head pose monitoring, open or closed eye detection. To make the drowsiness detection system automatic or practical for vehicles, the approaches used must be very fast. The facial information based methods have to detect or track the face robustly. After the face acquisition is properly done, the final processing can be done for feature extraction and classification.

6.3 Feature Extraction and Classification

In the feature extraction process, it is assumed that the proper location of the face is available and the faces are tilt corrected. Thus, it is assumed that the face have in-plane rotation only. The multiresolution based entropy features are extracted from the facial images in the same way as discussed in Chapter 5. Two methods have been followed to verify how well the multiresolution features can discriminate between drowsy and non-drowsy faces. The schematic diagram of the methods are given in Figs. 6.1 and 6.2. In the first method, the multiresolution entropy features are extracted from the training images and dimension reduction is performed using PCA and LDA. The query image is projected to the reduced subspace and classified using a nearest neighbour classifier. The method is same as the one proposed in Chapter 5. In the other method, the multiresolution entropy features are subjected to PCA space only. The classification is carried out using an SVM.

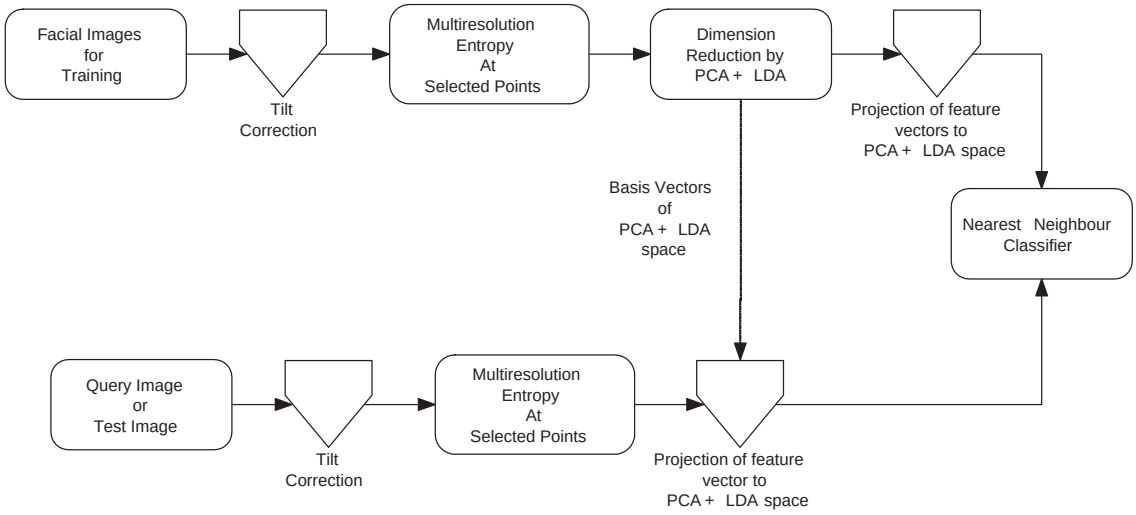


Figure 6.1: Drowsiness detection based on multiresolution entropy and nearest neighbour classifier

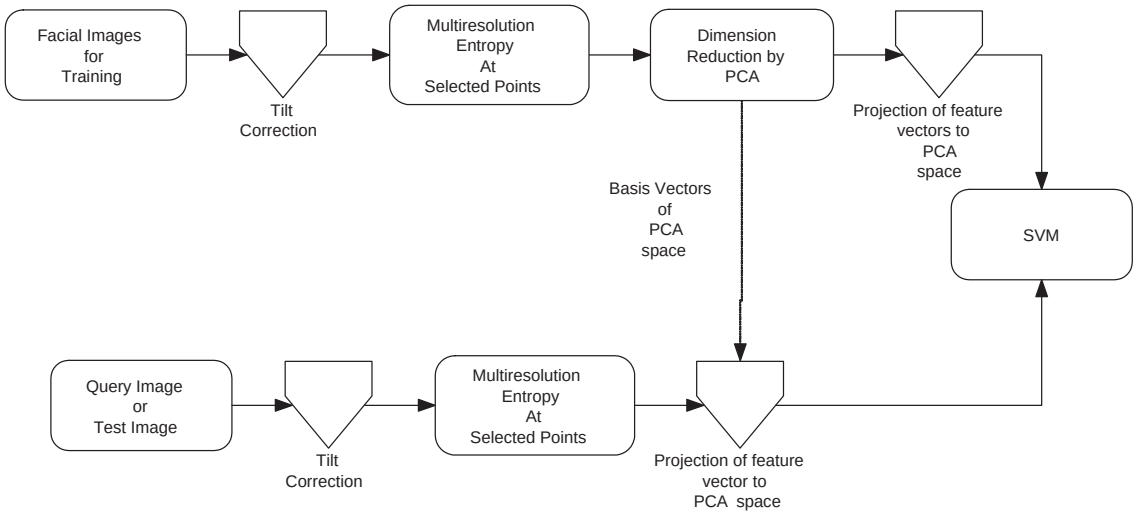


Figure 6.2: Drowsiness detection based on multiresolution entropy and SVM

6.4 Database

The previous researches carried out in this field have been evaluated in databases for private use only and they are not publicly available. Therefore, different methods have been evaluated using separate databases. Thus, a database has been generated for the drowsiness detection research of the thesis. The database consists of 8 videos from 8 University of Windsor students. There are seven male students and one female student. The ages of the subjects range between 26 to 35. The subjects in the database have different ethnicities; Indian, Chinese and Turkish. Also, the illumination conditions are different for each video. Apart from the drowsy faces in the videos, there also exist other expressions of the subjects. The other expressions from this database along with some expressions from the Cohn-Kanade database are used as negative images and drowsy faces are used as positive during training.

Figure 6.3: Some faces from the drowsy database

6.5 Results

The experiments are carried out with cross-validation only as drowsiness needs to be detected independent of person. The number of used images is 304. Due to cross-validated datasets, the test set has 20-30 images of 1-2 subjects and rest of the images are part of training set. Therefore, 5 test sets are formed. In separate experiments, curvelet based and contourlet based entropies at different scales and orientations are extracted. The scales and orientations used are same as those used in Chapter 5. Since the performance of wavelets are much lower than curvelets and contourlets, only directional multiresolution entropies are considered. The average recognition rate for different scales and orientations with varying number of principal components are shown in the Fig. 6.4. The maximum number of principal components used is 180 which is below the number of training images. The recognition rate at a specific orientation and scale for a multiresolution transform is the average recognition rate of the 5 sets. It is also found that, recognition rates given by curvelets are much better compared to that of contourlets. It is found that, with the increase of number of orientations, the recognition accuracy improves. The average recognition rates for the database using the first method are shown in Fig. 6.5. The maximum recognition rate of about 92% is achieved with curvelets.

Using the second method, with more scales and orientations, the recognition accuracy improves as shown in Fig. 6.6. The intermediate fluctuations in the recognition accuracy are also presented. The higher accuracies are given by initial principal components. This trend is exhibited in Fig. 6.4 as well. The maximum accuracy still remains above 90% here. The average accuracy in the database obtained using the second method is shown in Fig. 6.7. Again, the performance of curvelet entropy is better than contourlet entropy for drowsiness detection.

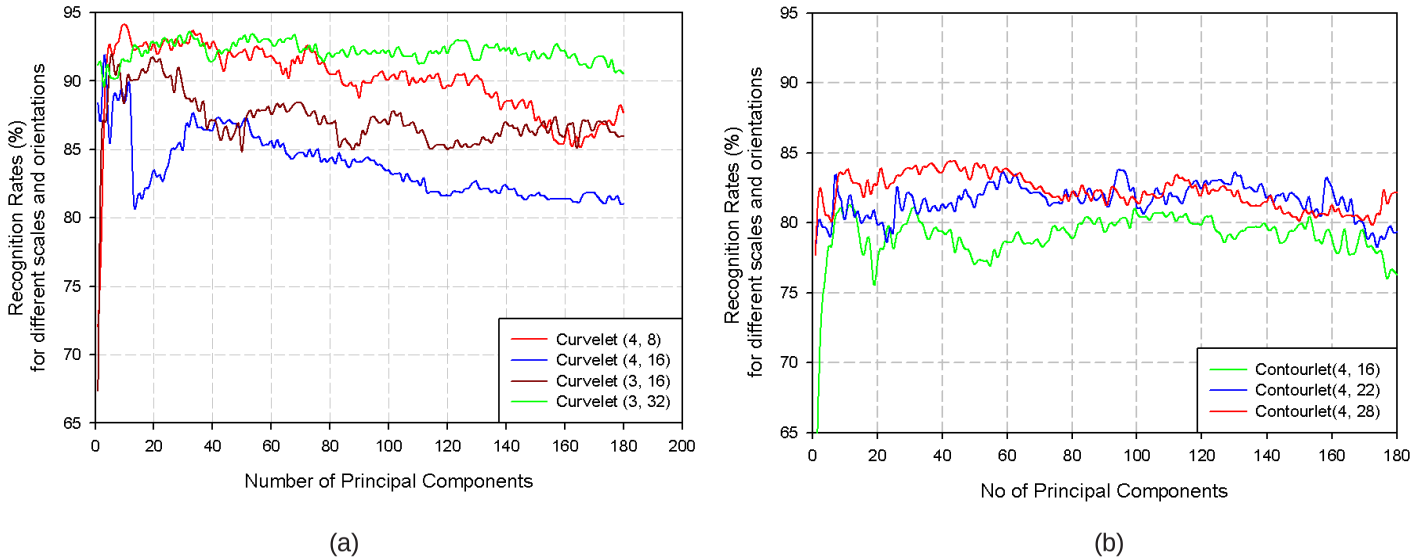


Figure 6.4: Recognition rates for drowsiness detection using nearest neighbour classifier with (a) curvelets (b) contourlets

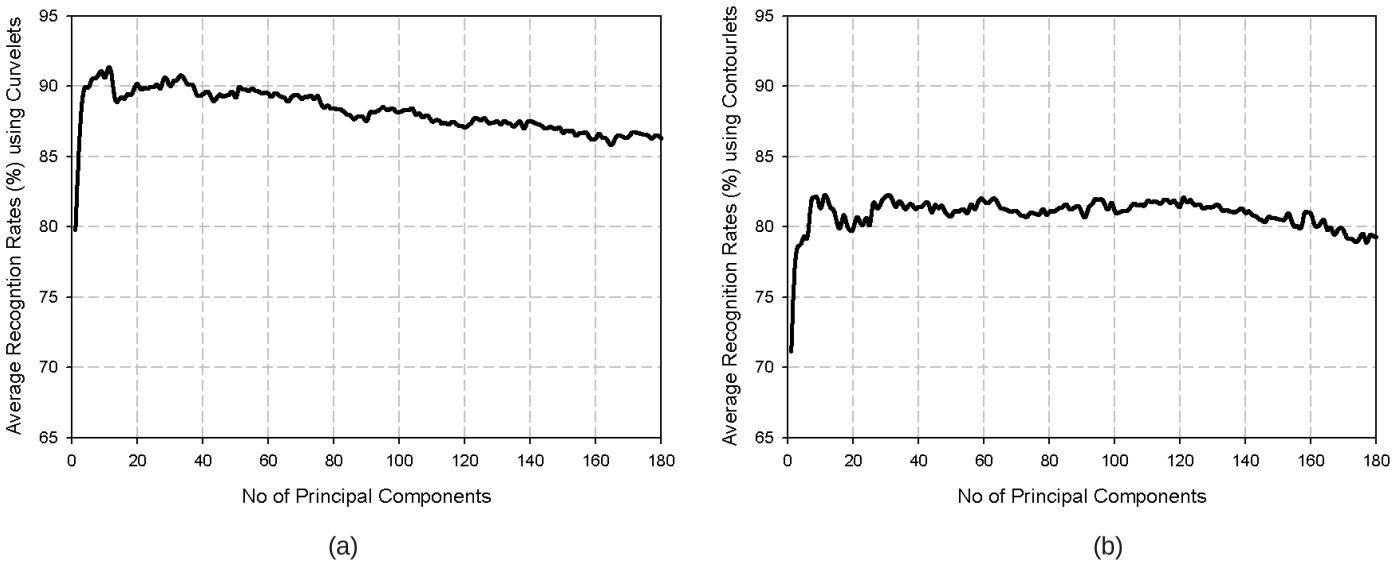


Figure 6.5: Average recognition rates for drowsiness detection using nearest neighbour classifier with (a) curvelets (b) contourlets

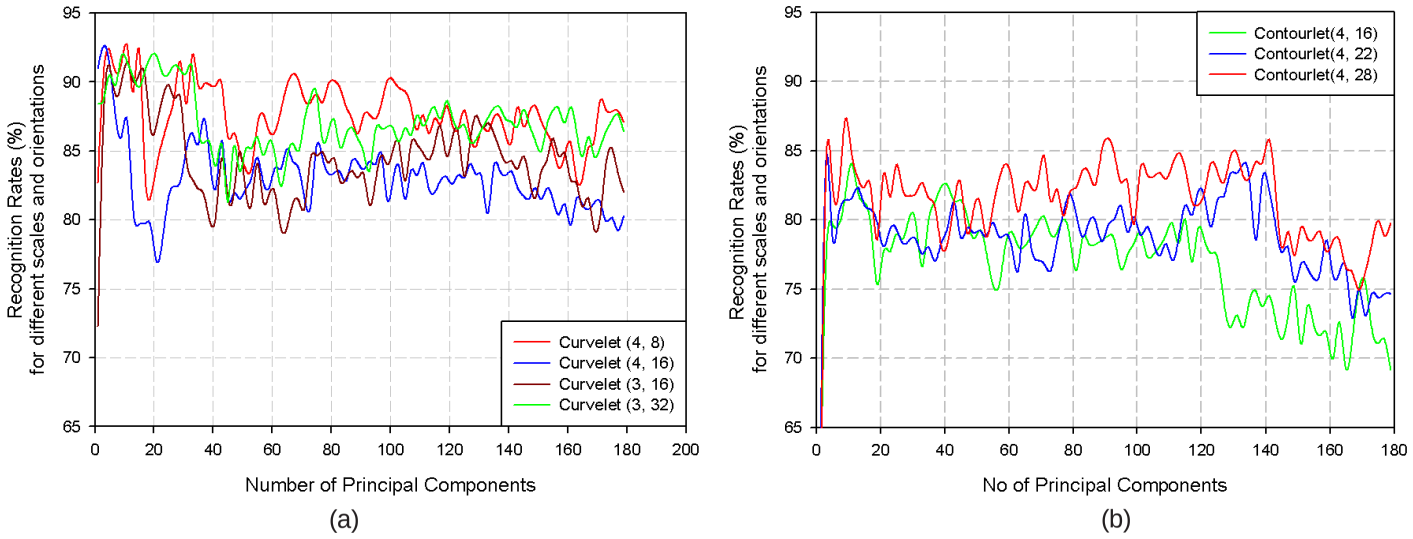


Figure 6.6: Recognition rates for drowsiness detection using SVM with(a) curvelets (b) contourlets

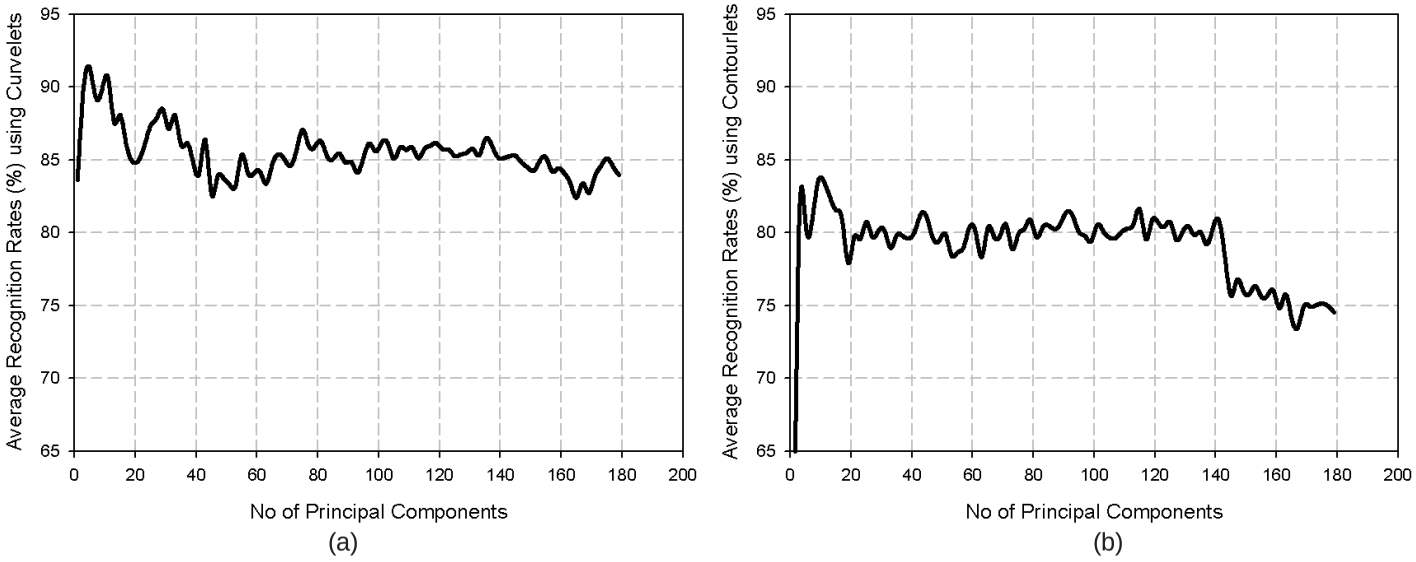


Figure 6.7: Average recognition rates for drowsiness detection using SVM with(a) curvelets (b) contourlets

6.6 Chapter Summary

The chapter has presented the use of multiresolution entropy for the purpose of drowsiness detection. Two methods using different dimension reduction techniques and classifiers have been used. The experiments have been conducted using a database generated for this purpose. The experiments are cross-validated and proves that directional multiresolution entropy is suitable for discriminating between drowsy and non-drowsy faces. Between the multiresolution methods used, curvelet entropy is found to give better accuracy than contourlet entropy.

Chapter 7

Conclusion

The thesis is concluded with a discussion of the research work and the scope for future work.

7.1 Contribution of the Research Work

The research work conducted in this thesis focusses on the use of directional multiresolution analysis for facial expression recognition. The orientation selectivity of directional multiresolution transforms has served as the motivation for the work. Two methods for facial expression recognition are proposed in two chapters. The penultimate chapter has focusses on a special case of facial expression recognition called drowsiness detection.

The first work proposed of the use of LBP, the popular texture analysis method in the curvelet domain. Experiments conducted on the JAFFE and Cohn-Kanade databases show that curvelet based LBP is better than LBP and Gabor wavelets. In this method, only the approximate subband of curvelet has been considered.

The second chapter proposed on the use of all curvelet subbands for facial expression recognition. To analyze different subbands, curvelet entropy at selected facial points has been used to form feature vectors. The dimension reduction is carried out using PCA and LDA. Also comparisons with contourlet based entropy and wavelet entropy for expression recognition has been provided. Non-cross-validated and cross-validated experiments have been performed on the databases for verifying the robustness of the approach. The proposed method was compared with Gabor wavelet

entropy and LBP. The results show that directional multiresolution transform like curvelets and contourlets can form effective features for expression recognition. These features are found robust against age, gender and illumination variations.

In the penultimate chapter, a study has been made to check the robustness of directional multiresolution entropy for drowsiness detection. Two classifiers– Nearest Neighbour and SVM have been used to classify dimension-reduced multiresolution entropy features. The experiments have been essentially cross-validated and curvelet entropy has provided a maximum accuracy of nearly 92% in discriminating between drowsy and non-drowsy faces.

As per the current literature, the use of directional multiresolution transforms for facial expression recognition is a novel idea. The results indicate that further scopes for research exist in this domain as discussed next.

7.2 Scope for Future Work

The experimental results of multiresolution based entropy exhibit fluctuations. A better idea may be to choose specific principal components by the method of boosting to achieve uniformity, and, hence better accuracy.

The method of multiresolution entropy works on selected facial points. If any of those points are occluded, the method faces problems. A scope therefore lies in re-engineering the method suitably for handling occlusions.

Finally, the work has used frontal or in-plane rotated facial images. To make the method suitable for different facial poses is a challenge and may be an area for further research.

References

[1] S. Bashyal and G. K. Venayagamoorthy. Recognition of facial expressions using gabor wavelets and learning vector quantization. *Engineering Applications of Artificial Intelligence*, 21(7):1056–1064, 2008. [7, 10, 11, 39]

[2] M. Black and Y. Yacoob. Recognizing facial expressions in image sequences using local parameterized models of image motion. *International Journal of Computer Vision*, 25(1):23–48, 1997. [7, 8]

[3] R. N. Bracewell. *The Fourier Transform and Its Applications (3rd ed.)*. McGraw-Hill, Boston, 2000. [13]

[4] P.J. Burt and E.H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, 1983. [16]

[5] E. J. Candès and D. L. Donoho. Curvelets, multiresolution representation and scaling laws. In *Wavelet Applications Signal Image Processing*, volume 4119, pages 1–12, 2000. [32]

[6] E.J. Candès, L. Demanet, D.L. Donoho, and L. Ying. Fast discrete curvelet transforms. In *Multiscale Modeling and Simulation*, volume 5, pages 861–899, 2006. [18]

[7] C. Chen and Z. Jhang. Wavelet energy entropy as a new feature extractor for face recognition. In *International Conference on Image and Graphics*, pages 616–619, 2007. [32]

[8] F. Chen, Z. Wang, Z. Xu, and J. Xiao. Facial expression recognition based on wavelet energy distribution feature and neural network ensemble. In *Global Conference on Intelligent Systems*, volume 2, pages 122–126, 2009. [32]

[9] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001. [10]

[10] M. Dailey and G. Cottrell. PCA gabor for expression recognition, 1999. Institution UCSD, CS-629. [10]

[11] C. Darwin. *The expression of the emotions in man and animals.* J. Murray, London, 1872. [2]

[12] M. N. Do and M. Vetterli. Pyramidal directional filter banks and curvelets. In *IEEE International Conference on Image Processing*, 2001. [19]

[13] M.N. Do and M. Vetterli. *Contourlets, Beyond Wavelets.* G. V. Welland ed., Academic Press, 2003. [19]

[14] D. L. Donoho and M.R. Duncan. Digital curvelet transform: Strategy, implementation and experiments. Technical report, Stanford University, 1999. [17]

[15] F. Dornaika and F. Davoine. Simultaneous facial action tracking and expression recognition using a particle filter. In *IEEE International Conference on Computer Vision*, pages 1733–1738, 2005. [10]

[16] G. Edwards, T. Cootes, and C. Taylor. Face recognition using active appearance models. In *European Conference on Computer Vision*, volume 2, pages 581–695, 1998. [8, 10]

[17] P. Eisert and B. Girod. Facial expression analysis for model-based coding of video sequences. In *Picture Coding Symposium*, pages 33–38, 1997. [9]

[18] P. Ekman and W.V. Friesen. Constant across cultures in face and emotions. *Journal of Personality and Social Psychology*, 17(2):124–129, 1971. [2]

[19] I. Essa and A. Pentland. Coding, analysis, interpretation and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):757–763, 1997. [7, 8, 9]

[20] I. A. Essa and A. P. Pentland. Facial expression recognition using a dynamic model and motion energy. In *International Conference on Computer Vision*, pages 360–367, 1995. [10]

[21] B. Fasel and J. Luettin. Automatic facial expression analysis: A survey. *Pattern Recognition*, 36:259–275, 1999. [6]

[22] W. Fellenz, J. Taylor, N. Tsapatsoulis, and S. Kollias. Comparing template-based, feature-based and supervised classification of facial expressions from static images. In *Circuits, Systems, Communications and Computers*, pages 5331–5336, 1999. [8]

[23] G.Zhao and M.Pietikäinen. Texture recognition using local binary patterns with an application to facial expressions. *IEEE IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):915–928, 2007. [10]

[24] H. Hong, H. Neven, and C. Vonder Malsburg. Online facial expression recognition based on personalized galleries. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 354–359, 1998. [7, 8]

[25] R. S. Huang, C.J. Kuo, L.L. Tsai, and O. T. C. Chen. Eeg pattern recognition arousal states detection and classification,. [58]

[26] Q. Ji and X. Yang. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance, 2002. [59]

[27] T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 46–53, 2000. [4]

[28] H. Kobayashi and F. Hara. Dynamic recognition of basic facial expressions by discrete-time recurrent neural network. In *International Joint Conference on Neural Networks*, page 155158, 1993. [11]

[29] A. Lanitis, C. Taylor, and T. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):743–756, 1997. [7, 10]

[30] S. Liao, W. Fan, C.S. Chung, and D.Y. Yeung. Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features. In *IEEE International Conference on Image Processing*, pages 665–668, 2006. [11, 21]

[31] J. Lien. Automatic recognition of facial expression using hidden markov models and estimation of expression intensity, 1998. Ph.D. Thesis, The Robotics Institute, Carnegie Mellon University. [9]

[32] C. Lisetti and D. Rumelhart. Facial expression recognition using a neural network. In *Flairs Conference, AAAI Press*, 1998. [10]

[33] M.J. Lyons, S. Akamatsu, M. Kamachi, and J. Goba. Coding facial expressions with gabor wavelets. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 200–205, 1998. [1, 8]

[34] S.G. Mallat. Wavelets for a vision. In *Proceedings of IEEE*, volume 84(4), pages 604–614, 1996. [15]

[35] T. Mandal, A. Majumdar, and Q.M.J. Wu. Face recognition by curvelet based feature extraction. In *International Conference on Image Analysis and Recognition, LNCS*, volume 4633, pages 806–817, 2007. [21]

[36] K. Mase and A. Pentland. Recognition of facial expression from optical flow. *The Institute of Electronics, Information and Communication Engineers Transactions*, E74(10):3474–3483, 1991. [6, 9]

[37] A. Mehrabian. Communication without words. *Psychology Today*, 2(4):53–56, 1968. [3]

[38] P. Michel and R. E. Kaliouby. Real time facial expression recognition in video using support vector machines. In *International Conference on Multimodal Interfaces*, pages 258–264, 2003. [11]

[39] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distribution. *Pattern Recognition*, 29 (1):51–59, 1996. [21, 22]

[40] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002. [22]

[41] T. Otsuka and J. Ohya. Spotting segments displaying facial expression from image sequences using hmm. In *International Conference on Automatic Face and Gesture Recognition*, page 442447. [11]

[42] W.V. Friesen P. Ekman. Facial action coding system: A technique for the measurement of facial movement. *Consulting Psychologists Press*, 1978. [3]

[43] M. Pantic and L.J.M. Rothkrantz. Expert system for automatic analysis of facial expression. *Image and Vision Computing*, 18(11):881–905, 2000. [8]

[44] M. Pantic and L.M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000. [6]

[45] M. Pantic and L.M. Rothkrantz. Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man, and Cybernetics*, 34(3):1449–1461, 2004. [10]

[46] C. Shan, S. Gong, and P.W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27(4):803–816, 2008. [7, 8, 10, 11, 21, 29, 47, 51]

[47] J. L. Starck. Image processing by the curvelet transform. PPT. [17]

[48] J.L. Starck, E. J. Candès, and D. L. Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6):670–684, 2000. [32]

[49] J. Steffens, E. Elagin, and H. Neven. Personspotter-fast and robust system for human detection, tracking and recognition. In *International Conference Automatic Face and Gesture Recognition*, pages 516–521, 1998. [7]

[50] M. Suwa, N. Sugie, and K. Fujimora. A preliminary note on pattern recognition of human emotional expression. In *International Joint Conference on Pattern Recognition*, pages 408–410, 1978. [6]

[51] T.Hamada, T.Ito, K.Adachi, T. Nakano, and S.Yamamoto. Detecting method for drivers drowsiness applicable to individual features. In *Intelligent Transportation Systems*, volume 2, pages 1405–1410, 2003. [59]

[52] T.Hayami, K. Matsunaga, K.Shidoji, and Y. Matsuki. Detecting drowsiness while driving by measuring eye movement - a pilot study. In *International Conference on Intelligent Transportation Systems*, pages 156–161, 2002. [59]

[53] Y. Tian. Evaluation of face resolution for expression analysis. In *Computer Vision and Pattern Recognition workshop on Face Processing in Video*, page 82, 2004. [8]

[54] Y. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):97–115, 2001. [9]

[55] M. Valstar and M. Pantic. Fully automatic facial action unit detection and temporal analysis. In *Computer Vision and Pattern Recognition Workshop*, page 149, 2006. [10]

[56] Paul Viola and Michael Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57:137–154, 2004. [7]

[57] A. Vuckovic, V. Radivojevic, A. C. N. Chen, and D. Popovic. Automatic recognition of alertness and drowsiness from eeg by an artificial neural network,. [58]

[58] W. Wierwille, S. Wreggit, R Fairbanks, and C. Kirn. Research on vehicle based driver status/ performance monitoring: development, validation and refinement of algorithms for detection of driver drowsiness, 1994. [59]

[59] Y. Yacoob and L.S. Davis. Recognizing human facial expressions from long image sequences using optical flow. *IEEE Transactions on Pattern Analysis and Machine Intellidgence*, 18(6):636–642, 1996. [9]

[60] P. Yang and Q. Liu aand D.N. Metaxas. Boosting encoded dynamic features for facial expression recognition. *Pattern Recognition Letters*, pages 132–139, 2009. [10, 11]

[61] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu. Comparison between geometry-based and gabor wavelets-based facial expression recognition using multi-layer perceptron. In *International Conference on Automatic Face and Gesture Recognition*, pages 454–459, 1998. [8, 10, 11, 33, 51]

Vita Auctoris

NAME	:	Ashirbani Saha
BIRTH YEAR	:	1983
BIRTH PLACE	:	INDIA
EDUCATION		
2008–2010	:	**Masters of Applied Science**
		Electrical and Computer Engineering
		University of Windsor, Windsor, Ontario, Canada
2002–2006	:	**Bachelors of Engineering**
		Electronics and Telecommunications Engineering
		Bengal Engineering and Science University, Howrah, India