**A THESIS REPORT**

**ON**

# PROPOSAL OF A UNIFIED ROBUST DESCRIPTOR FOR PATTERN IDENTIFICATION

**SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF**

## MASTER OF TECHNOLOGY

**IN**

## SIGNAL PROCESSING AND DIGITAL DESIGN

**SUBMITTED BY**


**PRACHI RAWAT**
**2K13/SPD/13**


**UNDER THE SUPERVISION OF**


**DINESH K VISHWAKARMA**
**ASSISTANT PROFESSOR**




**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING**

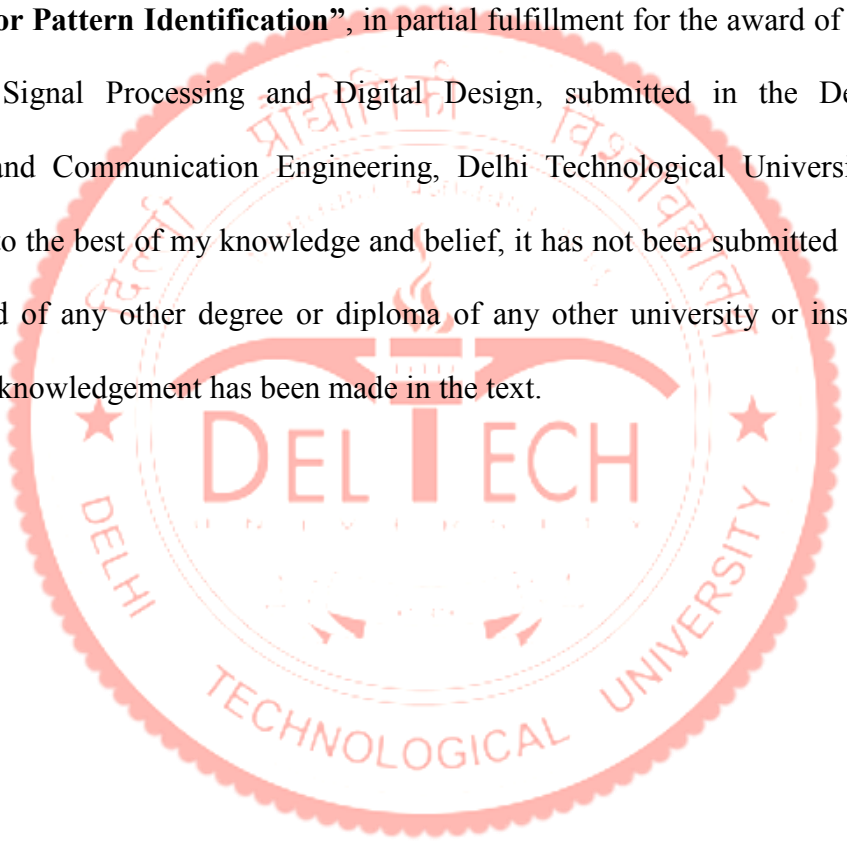
**DELHI TECHNOLOGICAL UNIVERSITY, DELHI**

**INDIA**


**JUNE 2015**

# DECLARATION

I hereby declare that the work presented in this report, titled **"Proposal of a Unified Robust Descriptor for Pattern Identification"**, in partial fulfillment for the award of the degree of M.Tech. in Signal Processing and Digital Design, submitted in the Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, is original and to the best of my knowledge and belief, it has not been submitted in part or full for the award of any other degree or diploma of any other university or institute, except where due acknowledgement has been made in the text.

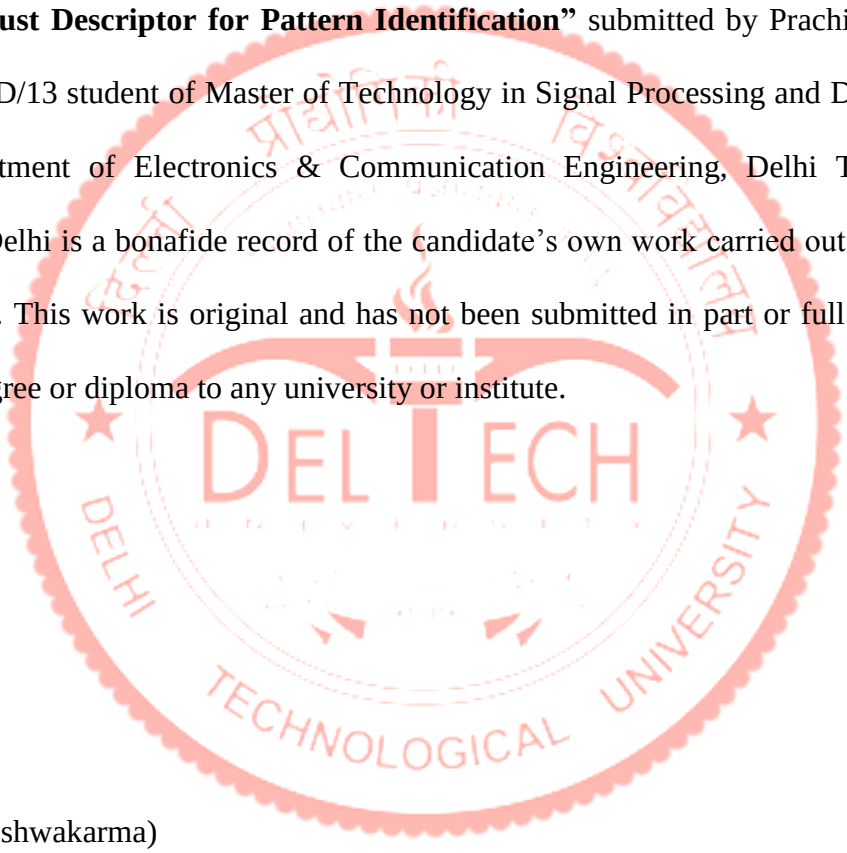



Prachi Rawat
Roll No. 2K13/SPD/13
M.Tech. Signal Processing and Digital Design

Date:

# CERTIFICATE

This is to certify that the research work embodied in this dissertation entitled **"Proposal of a Unified Robust Descriptor for Pattern Identification"** submitted by Prachi Rawat, Roll no. 2K13/SPD/13 student of Master of Technology in Signal Processing and Digital Design under Department of Electronics & Communication Engineering, Delhi Technological University, Delhi is a bonafide record of the candidate's own work carried out by her under my guidance. This work is original and has not been submitted in part or full for award of any other degree or diploma to any university or institute.




(Dinesh K Vishwakarma)
Assistant Professor
Department of Electronics & Communication Engineering
Delhi Technological University, Delhi, India

Date:

# ACKNOWLEDGEMENT

I express my deepest gratitude to my guide Dinesh K Vishwakarma, Assistant Professor, Department of Electronics and Communication Engineering, Delhi Technological University, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject. His suggestions and ways of summarizing the things made me go for independent studying and trying my best to get the maximum in my topic, this made my circle of knowledge very vast. I am highly thankful to him for guiding me in this project.

I am also grateful to Prof. Prem R Chadha, HOD, Electronics and Communication Engineering Department, DTU for his immense support.

Finally, I take this opportunity to extend my deep appreciation to my family, for their endless support during the crucial times of the completion of my project.

Prachi Rawat
Roll No. 2K13/SPD/13
M.Tech. Signal Processing and Digital Design

# ABSTRACT

In the field of computer vision, pattern identification plays a major role in the development of the system. To identify the patterns, there is a need of effective feature descriptor that describes and represents the characteristics of the object. The aim of this thesis is to develop a unified global descriptor using texture and shape information of the pattern. The texture information is extracted using Gabor Wavelet Transform (GWT) by developing a Gabor filter bank employing different orientation and scale. The shape information is obtained by applying Radon Transform on the image and the features are obtained by applying Discrete Wavelet Transform (DWT) on each of the Radon's projections. Since the DWT is being employed, the concept of discrete Radon transform is introduced in which the number of projections is dependent on the size of the image. By assimilation of Gabor feature with Radon feature, a unified descriptor is constructed and the performance of descriptor is evaluated by conducting an experiment using standard data sets. The performance is measured in terms of Average Recognition Rate (ARR) and further for its effectiveness the comparison of ARR is carried out with the similar state-of-the-art. The comparison of ARR shows the effectiveness and superiority of the proposed descriptor. The classification accuracy is evaluated through two widely used classification models, i.e. k-Nearest Neighbour (KNN) and Support Vector Machine (SVM).

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| 2D | Two Dimensional |
| 3D | Three Dimensional |
| ARR | Average Recognition Rate |
| CWT | Continuous Wavelet Transform |
| DFT | Discrete Fourier Transform |
| DOG | Difference of Gaussian |
| DWT | Discrete Wavelet Transform |
| GWT | Gabor Wavelet Transform |
| HOG | Histogram of Oriented Gradient |
| HMI | Human Machine Interaction |
| KNN | K-Nearest Neighbor |
| LKT | Lucas Kanade Tomasi |
| LOG | Laplace of Gaussian |
| MEI | Motion Energy Images |
| MHI | Motion History Images |
| MRI | Magnetic Resonance Imaging |
| RGB | Red Green Blue |
| SIFT | Scale Invariant Feature Transform |
| SURF | Speeded Up Robust Feature |
| SVM | Support Vector Machine |

# CHAPTER 1
# INTRODUCTION

The recent development in the field of artificial intelligence leads to reduction as well as replacement of human labor for the purpose of supervision. The field of computer vision deals with the methodologies related to artificial systems that derive information from images. The image data as an input to the computer vision system can have different forms, such as video sequences, multiple image views from different camera and multi-dimensional data from a medical scanner eg. MRI, ultrasonography and X-rays. The field of computer vision includes image reconstruction, image restoration, event or change detection, object recognition or pattern identification and tracking, object pose estimation, classification of objects and motion estimation.

In computer vision an object is represented by features obtained from important characteristics of an object in an image or a video. The performance of the pattern identification system relies on the algorithm used for representing the object by extracting appropriate features from it. Feature descriptors used in pattern recognition system are divided into two broad classes, global and descriptors. Global descriptor generalized the whole object of interest by a single vector while a local descriptor represents multiple patches of the image or a video sequence and is computed on multiple points present in it. Global descriptor computes appearance feature, texture feature, contour representation and shape based feature. Local descriptor includes Histogram of oriented gradient (HOG) features, Scale-invariant feature transform (SIFT), Speeded up robust features (SURF), Discrete fourier transform (DFT) and DWT. Global features require standard classification techniques such as SVM, KNN, decision tree and Naïve Bayes. Local features require specialized classification algorithm such as feature matching, Hausdorff average and maximum likelihood classifier.

The field of computer vision has much major application such as security system like face recognition, iris recognition, digital fingerprint and digital watermarking; as surveillance system at public places like airport, railway station, hospital and bank for monitoring abnormal behavior of any person; as machine vision in industry for automatic inspection and process control; as traffic management system by deployment of intelligent cameras for traffic laws enforcement by automatic number plate recognition, traffic sign detection and pedestrian detection; remote sensing for weather prediction; and development of human computer interface systems.

The thesis proposed a unified descriptor using Gabor filter and Radon transform. Features extracted from the Gabor filter and Radon transform are obtained by the application of DWT of the filtered or transform the image. The multi-resolution property of DWT is exploited in this work and is described later in detail. The descriptor is tested on the different poses of the human body to determine its accuracy.

## 1.1 MOTIVATION

The developing technologies in the field of computer vision, motivates to introduce a new descriptor that overcomes the disadvantages of the other descriptors. The global and local descriptor requires uniform brightness and illumination which is not possible in the practical scenario. The contour or shape based feature requires operations like segmentation or edge detection of the region of interest. The method develops here does not require such operations. The classifiers used for local descriptors, i.e. feature matching or maximum likelihood classifier is more complex than the standard classifier. The descriptor proposed using Gabor and Radon can be classified using standard classifiers like SVM or KNN. The descriptor works on the grayscale images, thus does not require any depth information and thus the system may not require installation of costly depth cameras. The descriptor can be used for recognition and classification of any object. In this thesis the descriptor is tested on human body poses and thus a complete system is determined for human action recognition in a video sequence.

## 1.2 CHALLENGES

There are many challenges that came across while defining the descriptor as well as the system for human activity recognition.

- The descriptor must consist of characteristics like scale, translation and rotation invariance. Many descriptors fulfill one of these characteristics at a much higher rate than compared to other characteristics. SIFT features focus on the scale invariant properties, while HOG focuses on translation and rotation invariant properties.

- The local descriptors consist of many vectors and a proper methodology is to be defined to obtain a feature vector for image representation. The local descriptor basically gives two vectors, one is the key point location and the other is feature descriptors. It provides a large no of key points which may be dense at the same location of the object while rare in some other location. A proper methodology is to be developed to represent a feature vector of the same length that would be sufficient to represent all different objects.

- The local descriptor also requires complex algorithms for classification such as feature matching, hausdorff average and maximum likelihood classifier, thus making the system complex. A descriptor should be defined such that it can be easily classified using standard classifier using SVM and KNN.

- The global descriptors based on appearance require uniform illumination and brightness for all the images out of which features are calculated. This condition is not fulfilled in practical situations as there are temporal variation in illumination and brightness. Thus a descriptor must be invariant to brightness and illumination.

- Occlusion of the object is also a major challenge for recognition. The feature based on the presence of the whole object does not work properly when the object is concluded. Occlusion does not affect the performance of the descriptors obtained from key point features as it computers the features on a small part of the object.

- Segmentation is a method to obtain the region of interest and usually an initial step before extracting features. Segmentation consist is a very complex algorithm and it cannot be uniformly applied on different images with different background or illumination. As a global descriptor based on contour requires segmentation makes the system less efficient.

- Viewpoint based observation also creates problems in predicting the shape of the object. This challenge plays a major role in human pose recognition. A surveillance system consisting of a large number of cameras may consist of different viewpoint of the human body, thus proper processing is required to perform operations on the correct viewpoint pose.

- Proper key poses determination from a video sequence for training the system is very important to develop a robust system for human activity classification. A lot of data samples obtained from the video sequence would be redundant and must be removed. In this way better inter class classification is obtained. In many cases, intra-class motion variability also exists, i.e. same person can perform a single action in a different way.

## 1.3 APPLICATIONS

The major applications of the pattern identification system in the era of artificial intelligence are described below:

- **Security Systems**

In the era of artificial intelligence security is made reliable on the autonomous systems. Techniques developed for iris recognition, face recognition and thumb recognition are some example of it. All these security system requires a high end scanner and a robust image recognition system for maintaining high accuracy.

- **Video Surveillance**

Video surveillance is deployed in public places like railway station, airport, bus stop, banks and shopping complex; organizations like offices, school, colleges and hospitals for

keeping a check on the activity of the human being for the purpose of security, as safety is the outmost important for any organization and for which they install a large number of cameras. Manual monitoring of the videos from these cameras is not possible, thus it requires an intelligent system to automatically determine any abnormal human activity

- **Remote sensing**

Satellite image processing is one of the key research area in determining data such as weather forecast, analysis of biomass content, analysis of moisture content, mapping of thermal profile, identification of mineral deposit, to review the type of vegetation and peak vegetation, for the analysis of deforestation of amazon basin, monitoring of glacial features, the cloud and surface mapping, land-water bounds, recognition of snow and ice, sea surface temperatures and cloud cover analysis at night.

- **Monitoring of an Individual**

Monitoring an elderly person or child at home, an athlete in the practicing area or a patient in a room using an intelligent system we can provide them with immediate care and attention. As the system can be developed such that it can send information to the concerned person in case of elder or child and to doctor in case of patient or athlete. These systems are increasing day by day for the safety of an individual.

- **Gesture Recognition**

Gesture recognition is developed for human machine interaction (HMI) that a computer or a machine can understand the actions of the human without any manual or text input. It includes the analysis of the human hand or body or the facial expression. One of the major application is the analysis of the sign language by a computer to determine what a person wants to say and becomes a mode of communication such that other people can understand.

- **Traffic enforcement system**

The automatic number plate recognition can lead to enforcement of traffic laws on roads an highway. The system can study the pattern of the movement of the vehicles on the roads

and can recognize the vehicles not following the rules and can store the information about the vehicle along with the registration number of the vehicle.

- **Intelligent character recognition**

In this a system is designed such that a computer can learn the different styles of writing for increasing accuracy and rate of recognition. It is also known as optical character recognition. It creates a system that could analyze the characters in any scanned image document and can create an actual document file of it in the system.

- **Machine Vision**

It provides an image processing based automatic inspection, product quality analysis, product packing analysis, raw material analysis, product serial no analysis and robot monitoring in the industry.

## 1.4 LAYOUT OF THE THESIS

The thesis content is as follows:

Chapter 2 gives the literature review of the different descriptors used for extracting features for the application of pattern identification. It describes features into two broad categories, local features and global features. Different types of local features and global features are explained along with its application and usage.

Chapter 3 describes the proposed methodology in detail. It explains the concept of using a descriptor that combines different features gives robustness to the system. It explains the concept of wavelet transform, Gabor function, GWT and Radon transform. It also explains the process for extracting the features from the GWT as well as Radon wavelet transform.

Chapter 4 gives the experimentation and the result of the methodology proposed. The feature descriptor is applied of different type of dataset including both video frames and

images. The data sets were hand gesture data sets such as Jochen Triech, NUS I and Cambridge as well as human activity data sets such as KTH and Weizmann.

Chapter 5 gives the conclusion and future scope of the methodology proposed in this thesis. Human activity and hand gesture recognition are the two applications for which the accuracy of the descriptor is determined. The descriptor can be further verified for other different applications in the field of computer vision.

# CHAPTER 2
# LITERATURE REVIEW

On the basis of the previous work done in the field of pattern recognition, the descriptors are divided into two broad classes as local descriptor and global descriptor. Global descriptor generalized the whole object of interest by a single vector while a local descriptor represents multiple patches of the image and is computed on multiple points in the image. The accuracy of the recognition system depends on the features extracted from an image as well as the methodology used for classification.

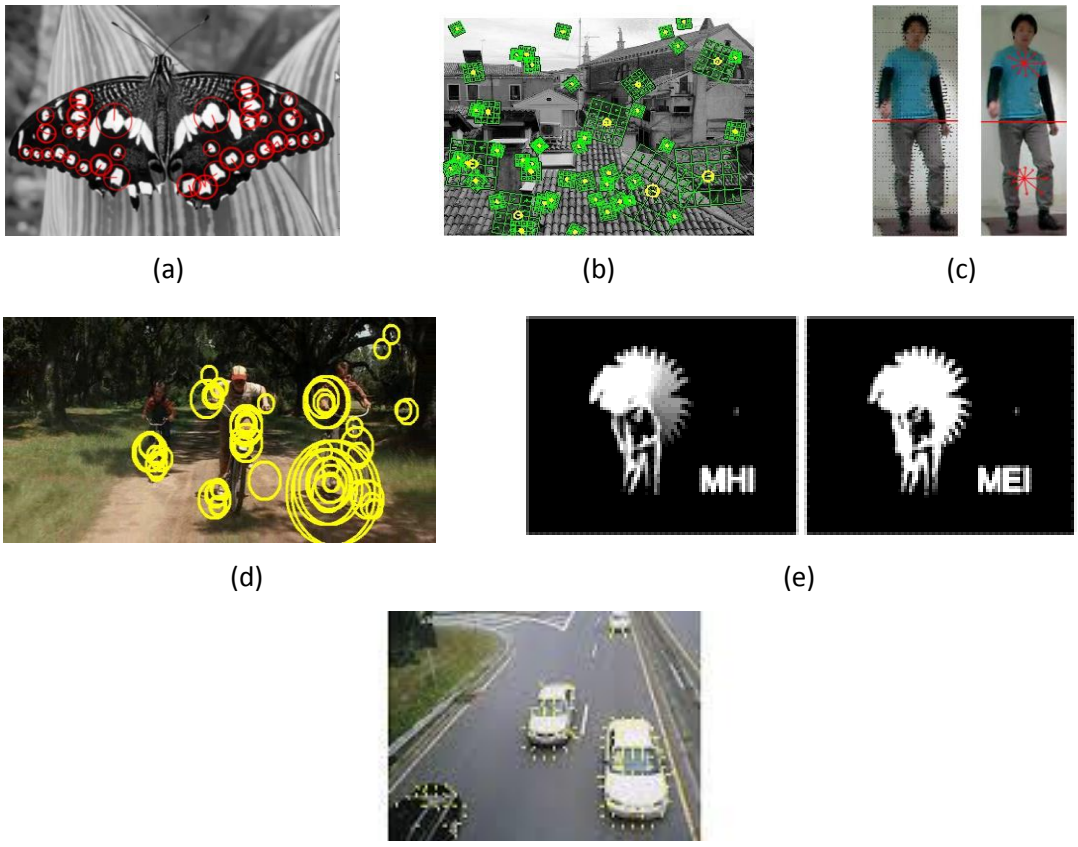
## 2.1 LOCAL DESCRIPTOR

Local descriptors are obtained at the local image neighborhood around the interest point also known as a key point. The key points are detected at multiple scales and are selected such that they are invariant to different views of the object. The feature descriptor gives the information of the image patch around the key point. For a video sequence the object in motion can be detected by analyzing the same patch in consequent frames. The features, detecting the motion can be used to track the pattern of the movement of the object. Local descriptor includes Speeded up robust features (SURF), the Scale-invariant feature transform (SIFT), Histogram of oriented gradient (HOG), Harris3D detectors, Cuboid detectors and Hessian detectors. Local features require specialized classification algorithm such as feature matching, hausdorff average and maximum likelihood classifier.

In 2006, Herbert Bay et al. [1] proposed Speeded up robust features (SURF) which is robust as compared to other descriptors. The name is such that it can be computed and compared faster than other features. SURF descriptors are invariant to scale changes, image rotation, viewpoint changes and illumination change. It performs second order Gaussian differentiation over the image and a Hessian matrix is obtained. The key point location and

scale value are obtained by the determinant of the Hessian matrix. In 2008, Kim et al. [2] proposed SURF features for face component recognition.

In 1999, David G Lowe proposed a methodology for pattern recognition using scale invariant feature Transform [3] [4]. SIFT [5] descriptors are invariant to scale, translation and rotation. The features are obtained using two methods, namely Laplace of Gaussian (LOG) and Difference of Gaussian (DOG). LOG is less efficient and costly method, therefore DOG is the more popular method. In this different image octaves are subtracted at different scales to obtain a key point. In 2008, C. Fernandez and M.A. Vicente [6] proposed face recognition methodology using SIFT features.




(a) (b) (c)

(d) (e)

(f)

Figure 2.1 Local descriptors (a) SURF, (b) SIFT, (c) HOG, (d) Harris 3D, (e) MHI & MEI, (f) LKT

In 2005, Dalal and Tiggs [7] introduced histogram of oriented gradient (HOG) for identification of human present in an image. The HOG is obtained by determining normalized local histogram of the gradient orientation values of the image. HOG features are prominently used in the detection of the presence of humans in an image. In 2013, Sebastian Tuermer et al. [8]Usedd HOG feature for the determination of air bornee vehicle and disparity maps. In 2014, S. Gunasekar et al. [9] proposed the methodology face detection using HOG features.

When we are considering a pattern identification problem for tracking an object motion in a video we consider the features that consist of temporal information also. In 2003, Laptev and Lindeberg [10] introduced Harris3D detector as a space-time feature. It calculates the second moment at each frame point using independent scale values and gradient values at a particular space and time. The final space-time interest point location is given by the local maximum values over a given region. In 2009, Guo Chenguang et al [11] proposed the methodology for fast computation of corner detection using Harris detectors. It uses Gaussian smoothing link to increase the performance of the Harris detectors. In 2010, Jiguang Dai [12] proposed Harris detectors for matching of remote sensing images. It combines Harris detectors with SIFT keypoints for robust detection. In 2011, J.B. Ryu et al. [13] proposed the methodology for shape based corner classification using Harris detectors. The algorithm was implemented on basic geometrical shapes.

In 2005, Dollar et al. [14] proposed Cuboid detector which is based on the Gaussian smoothing kernel. The interest points are concluded by the local maxima obtained among the result of smoothing function. In 2008, Willems et al [15] proposed Hessian detector as a space-time measurement of the Hessian saliency measurement used for detecting the moving object in the image [16] [17]. The determinant of the Hessian matrix is obtained as the saliency value. The determinant of the hessian is obtained over different scales of space and time. The feature point is detected by the non-maximum suppression algorithm.

In 2001, Bobick and Davis [18] introduced the concept of temporal templates for detecting the motion of a human body and constructed MHI/MEI images from consequents

video sequences. These templates are scalar vectored images over time sequence and thus a function of motion and as well as time. The main drawback of these templates was its view dependent properties. Different templates are obtained from a different view point. If the speed of the action is high, then also the template cannot be obtained with accuracy.

In 2005, Sheikh et al. [19] models the motion of a human body using space time trajectories. It tracks the motion of 13 human joints over consecutive frames and obtains a normalize space time trajectories. In 2004, Liu et al. [20] proposed the methodology of LKT [21] [22] features to extract the motion feature in a video. In this the motion features are extracted by tracking the points in the key frames and the original frames. The motion is considered to be localized, i.e. LKT feature values are similar in nearby points. The drawback of this feature is in the case of large motion in a video the method does not calculate the motion feature accurately

## 2.2 GLOBAL DESCRIPTOR

Global descriptor computes appearance feature, texture feature, contour feature and shape based feature. Such features are of great significance as the whole image is represented by a single point in a high dimensional space. Thus, standard classifiers such as SVM or KNN can be used for classification. Appearance based feature extracts the colour information in the image.

Swain and Ballard [23] in 1991 proposed colour indexing which recognize the object by colour histogram indexing. The number of bins is finite and represents a color in RGB space. Stricker and Orengo [24] in 1995, proposed colour histogram in cumulated form, giving much better results than previous one. Rui and Huang [25] in 1997 introduced colour correlogram which defines the spatial correlation between the pair of color with distance in an image. In 2000, Liu and Xiong Mean [26] of the colour values is calculated at different regions. The contour features consist of mathematical representation of the edges of the image. The edge direction is computed as a histogram of direction or as density of edges

[27].The shape based features are generally geometric features like Circularity, Spreadness, Roundness, Solidity, Eccentricity, Convexity and Irregularity as described by Dimitri et. al [28].

Contour feature can be obtained by different methods such as a normalized count of the white or black pixel in the segmented video sequence. In 2011, Rahman et al [29] [30] proposed a methodology of analyzing white spaces in the segmented video sequence to determine the activity present in the video sequence. In 2008, Meghna et al [31] encodes the information about the edges of the silhouette by using a chain code methodology. In chain code a number is assigned at each pixel value of the edge according to the location of the next pixel value. Nergui et al [32] obtained contour features from the silhouette of the segmented image. The features obtained were the height to weight of the contour data, the centroid point of the upper past and lower part and the distance between the two centroids. . In 2010, Venkatesha and Turk [33] describes a method of obtaining feature points on the contour edge image of the human body in a given video sequence. The feature points are located on the edges of the contour and a minimum of 30 points was obtained. The local features that were extracted from the key points where the turning angle and the distance across the shape.

In 2011, Kumari and Mitra [34] proposed DFT for the analysis of the image over a small area for determining the frequency characteristics in the image. It is assumed that there is a difference bin the gray level intensity of the foreground object image and the background image. Thus the feature vectors can be drawn from the sub image representing the shape or the geometric structure of the object. In 1910, wavelet transform was first introduced by Haar, but the major development took place in 1980 by the work of Morelet. In 1999 S. Pittner and S. V. Kamarthi [35] proposed the feature extraction methodology for pattern recognition. DWT can be used to extract shape features of an object in the image. Energy, Entrophy and function approximation are the most important features that can be obtained from an image after application of DWT.

In 1946, Dennis Gabor introduced Gabor filter and then Daugmann [36] applied a Gabor filter to image texture in 1980. The energy content of the filtered image at different scale and orientation is obtained in determining the features. The energy at a particular scale and orientation is defined as the summation of all the GWT coefficients at that particular scale and orientation. A conclusion is obtained that the higher energy determines the highest information content at that scale and orientation. In 1992, Bovik et al [37] proposed the methodology to use Gabor filter for estimating the emergent image frequency present in the image. The assumption taken image consist of a single concentration on emergent frequencies in a local area. Tai Sing Lee in 1996 [38] described the methodology to represent the image using 2D Gabor wavelets using Daubechies's wavelet. It describes the methodology to obtain a filter bank for Gabor wavelets and to determine the frame bound for it. Lee concludes that a tight frame represents high resolution images as it performs experiments with severely quantized Gabor coefficient. In 2008, Wei Jiang et al. [39] used Gabor wavelets to detect the edges in the image and proves the methodology as highly efficient in terms of detection accuracy and computational accuracy. In 2006, S. Arivazhagan et al. [40] describes a methodology for analysis of different patterns, i.e. texture using Gabor wavelet. The features obtained to determine the rotation invariant property of the Gabor wavelet. Other approaches that described the use of Gabor in texture estimation are given in [41-45].

Radon transform is widely used in the field of image processing such as image retrieval, denoising of image, image watermarking, image reconstruction, line detection, image restoration and image hashing. In 1998, Mostafa A. Ibrahim et al [46], proposed the methodology of using Radon domain. In 2007, An Zhiyong et al [47] proposed the methodology for content based image retrieval based on wavelet transform and Radon transform. In this the edges and lines of the images are obtained at different scale by using wavelet transform and then Radon invariant features are obtained by the multiscale edge image.

In 2011, Z. A. Khan and Won Sohn [48] described R-Transform is also a mathematical parameter derived from Radon transform that represents a shape of the object. R-Transform

being scaled and translation invariant is a very important descriptor. In 2006, S. Arivazhagan et al [49] proposed a methodology to extract feature for texture classification and also describes its rotation invariant property. The features extracted from the image obtained by applying wavelet transform on Radon image, as introduced by S. Arivanzhagan are energy, contrast and homogeneity, which were derived from the sub-bands of the wavelet decomposition. The co-occurrence matrix for each of the decomposition is also one of the most important features for classification. In 2010, Satyajit Kautkar et al. [50] uses Radon and wavelet transform for face recognition and it obtains the features from the LL part of the wavelet decomposed face image.

# CHAPTER 3
# METHODOLOGY

The thesis describes a new methodology for pattern identification by defining a new unified descriptor based upon the DWT. The features are obtained by combing texture feature and shape feature. These feature descriptors obtained are invariant to changes in scale, translation and rotation and hence the descriptor is robust. In 2007, D. Zang et al [51] proposed a method to describe local invariant features combing luminance and colour information. In 2009, Pedram Azad et al, [52] combined Harris point detectors and SIFT descriptor to obtain scale invariant pattern recognition. In 2012, Ying Li Tian [53] proposed a combination of spatial and temporal descriptor by using HOG and MHI methodology for determining the features. In our methodology the GWT is used to extract the texture information in the image and the information thus obtained is invariant to scale and rotation. The Radon transform is also applied to the images to obtain the scale and translation invariant features. The Gabor wavelet performs multi resolution analysis of the image at different scale and orientation. The DWT is applied to each of the Radon projections for obtaining the robust feature. The Fig. 3.1 describes the detailed layout the methodology described in this thesis.

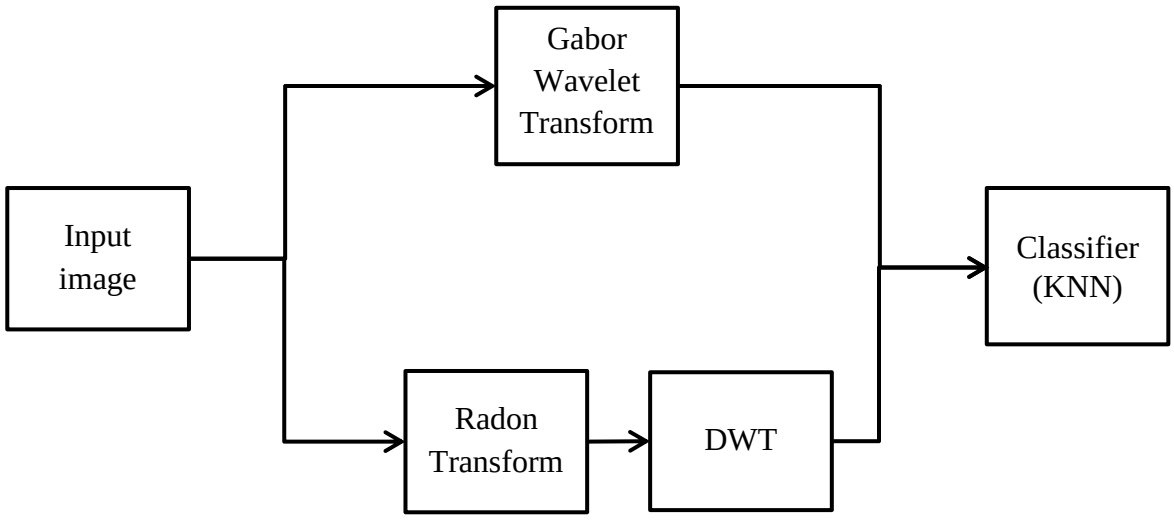



Figure 3.1 Overview of the proposed methodology

Many applications of this descriptor are also proposed in the thesis. Different types of feature can be obtained from the result of Gabor wavelet filter and Radon transform of the image.

## 3.1 WAVELET

Wavelet is the powerful mathematical technique employed in the field of pattern recognition. In the field of image processing it has initially been used for compression of image eg. JPEG2000. In 1910, the first wavelet Haar families were introduced, but the concept of wavelet was much developed in 1980 with the theory given by Morelet. In 1988 Daubechies's [54] explained the systematical method for orthogonal wavelets and Stephane Mallat gives the concept of multi-resolution analysis.

Wavelet can be used to separate the different scale components of the image and a frequency can be assigned to it. A resolution that matches with the scale component is considered and represented by the wavelets. Wavelet is considered as a set of orthogonal oscillating function which is scaled and translated at different values. Wavelet consist of many pattern recognition applications such as shape descriptor [55], singularity detector [56], invariant representation [57], character recognition [58], texture processing [40], human face recognition [50], document analysis, character font design and noise removal [59]. The wavelet transform is divided based on the type of the data on which it is operated, as the continuous wavelet transform (CWT) and DWT.

### 3.1.1 Continuous Wavelet Transform

A CWT can be operated at every possible scale and translation values. CWT can be applied on any analog signal or data. The expression for CWT for 1-dimensional signal is given by Eq. 3.1.

$$C_x(\alpha,\beta) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{\infty} x(t)\ \psi\left(\frac{t-\beta}{\alpha}\right) dt \tag{3.1}$$

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

Where $x(t)$ is the input signal, and $\psi(t)$ is the mother wavelet. The constant value $\beta$ determines the position or the location and $\alpha$ determines the scaling value of the function. Both $\alpha$ and $\beta$ are any positive real number.

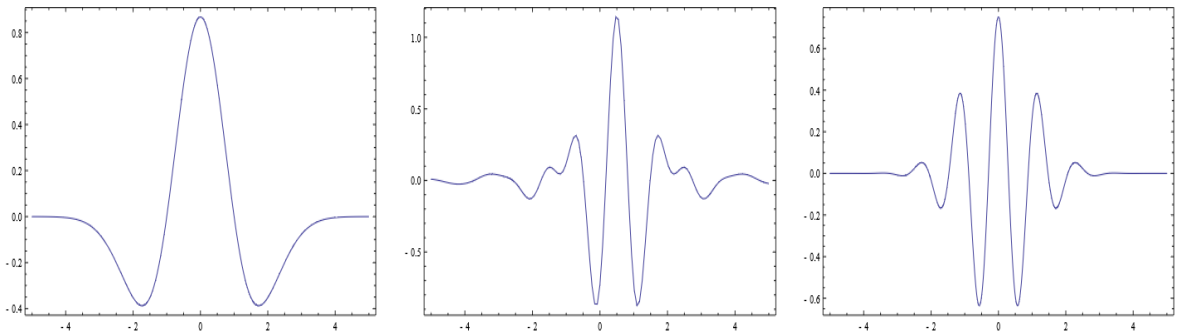
The scaling function is given by Eq. 3.3

$$\emptyset(t) = \int_{-\infty}^{\infty} \varphi(f)\ e^{j2\pi ft}\, df \tag{3.3}$$

Where $\varphi(f)$ is given by Eq 3.4

$$|\varphi(f)|^2 = \int_f^{\infty} \frac{|\psi(f_1)|^2}{|f_1|}\, df_1 \qquad \text{for} \quad f > 0\,, \quad \varphi(-f) = \varphi^*(f) \tag{3.4}$$

The mother wavelet is considered s high pass filter and the scaling function as low pass filter. Some example of the 1D wavelet function is shown in Fig. 3.2.



(a) (b) (c)

Figure 3.2 Mother wavelet function (a) Mexican Hat wavelet, (d) Mayer wavelet (c) Morlet wavelet

The scale factor $\alpha$ is used for the purpose of compressing or dilating the signal. If the value of the scale factor is small, then the signal is contracted which results in more detailed information in the resultant signal. Whereas if the value of the scale factor is larger the signal is elongated this results an output signal with less detail.

The equation for CWT in 2-dimension is given by Eq. 3.5 and a 2-D Mexican hat wavelet is described in Fig. 3.3.

$$C_r(\alpha,\beta,\gamma) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r(x,y)\ \psi\left(\frac{t-\beta}{\alpha}\right)\ \psi\left(\frac{t-\gamma}{\alpha}\right) dx\ dy \tag{3.5}$$

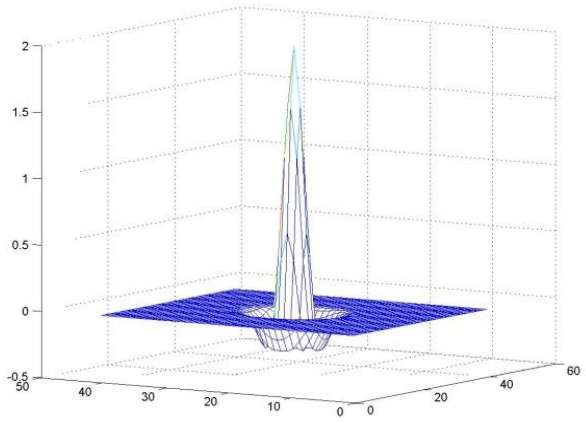



Figure 3.3 Mexican hat wavelet in 2D

The CWT is difficult to implement in a real scenario, thus a great advancement developed for DWT along with many applications.

### 3.1.2 Discrete Wavelet Transform

DWT is the discretized form of CWT. DWT of a wavelet is calculated by passing the signal through a quadrature filter bank that consist of a high pass filter and low pass filter as the wavelet function and scaling function respectively. The system is represented in Fig. 3.4.

In Fig. 3.4, x[n] is the input signal, g[n] is the high pass filter and h[n] is the low pass filter. The signal after passing through the filter is decimated by a factor of 2. The resulting output from the high pass filter gives the approximation coefficient and from the low pass filter gives the detailed coefficient. The approximation coefficients consist of less information, hence it is removed in the process of compression of a video or signal.

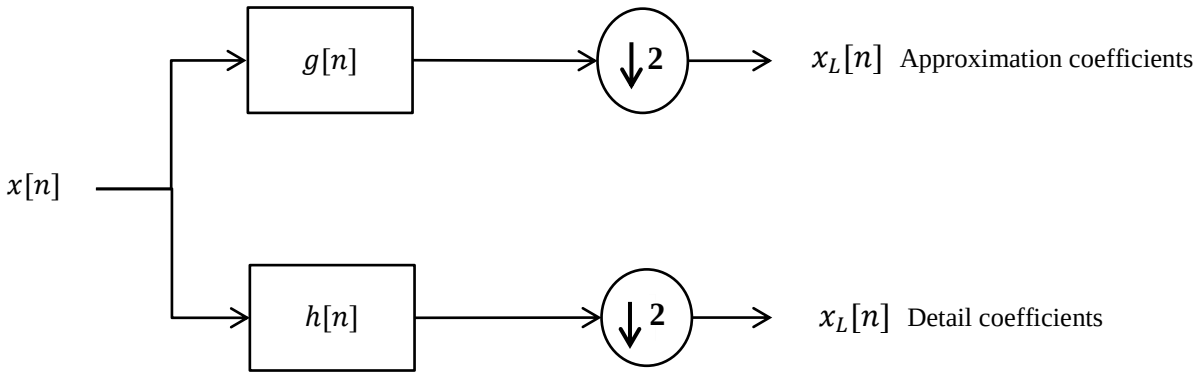



Figure 3.4 Discrete Wavelet Transform

### 3.1.3 2D Discrete Wavelet Transform

2D DWT is obtained by performing 1D DWT along both the dimensions one after another. The model is described in Fig. 3.5. 2D DWT is performed on images for compression operation and for determining edges or corners. The low pass filter output, $x_L[m,n]$ consist of the detailed coefficients and the majority of the image information lies in it. While the other outputs $x_{H1}[m,n], x_{H2}[m,n]\ and\ x_{H3}[m,n]$ consist of approximation coefficient and thus less information content is present in it. The figure represents the single stage DWT on a 2D data. The system can be cascaded to obtain the multistage DWT for a 2D data.

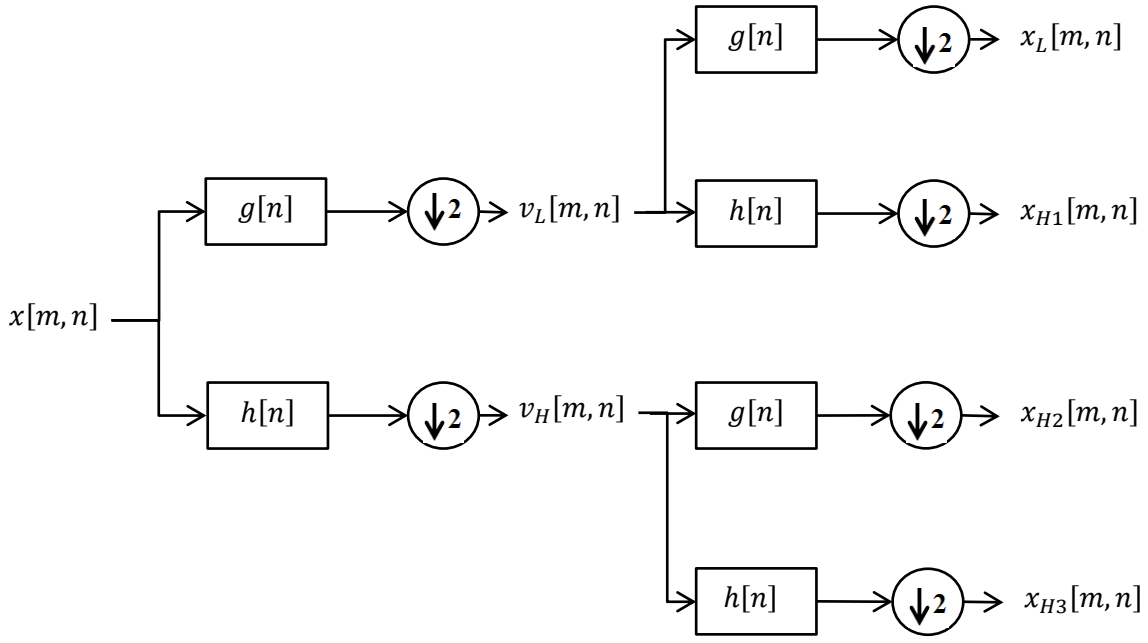


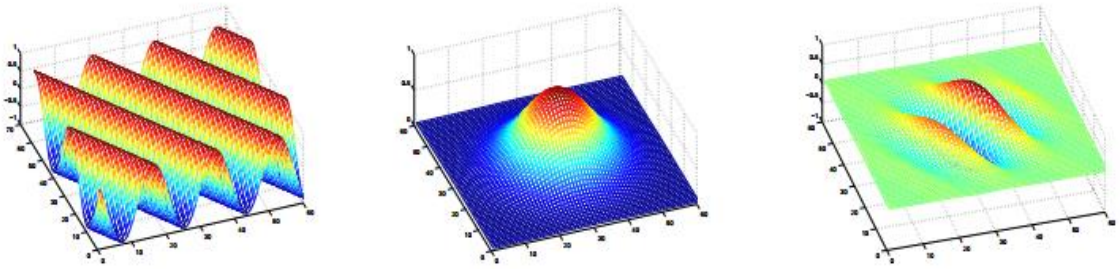

Figure 3.5 2D DWT

## 3.2 GABOR FUNCTION

In 1946, Dennis Gabor introduced the concept of Gabor function, in which a bank of filters is created by the translation and modulation of a generating function. Gabor function is obtained as the complex harmonic plane wave function modulated by a Gaussian function. Gaussian function is a linear band pass filter and is also considered as a non-orthogonal basis function. The general 1D form given by Eq. 3.6.

$$G(x) = \frac{1}{\sqrt{2\pi\sigma}} \; e^{\left(\frac{-x^2}{2\sigma^2}\right)} e^{(j\pi W x)} \tag{3.6}$$

The 2D Gabor function is given as modulation of a 2D plane wave with a 2D Gaussian function. Fig. 3.6 depicts the formation of a 2D Gabor filter.



(a) (b) (c)

Figure 3.6 Gabor filter construction, (a) 2D sinusoid with 30° with x axis, (b) Gaussian function, and (c) corresponding Gabor filter

The mathematical equation of a 2D Gabor function is given by Eq. 3.7.

$$G(x,y) = g_\sigma(x,y)\, e^{(2\pi j\omega(x\cos\theta + y\sin\theta))} \tag{3.7}$$

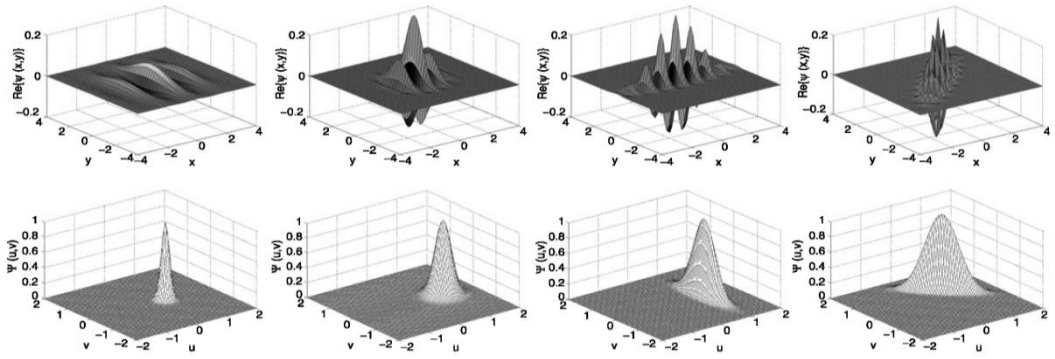
$$g_\sigma(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right) e^{\left(\frac{-1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)\right)} \tag{3.8}$$

In Eq. 3.7, $g_\sigma(x,y)$ is the 2D Gaussian function and its mathematical formula is given by Eq. 3.8. The key parameters of the Gabor functions are the oscillating frequency W, the scale parameter $\sigma_x$ and $\sigma_y$, in x and y direction and the orientation parameter $\theta$. A bank of filters is obtained by tuning the scale and orientation of the function. The 2D function performs optimal localization in both spatial and frequency domain. The 2D filters have been widely used in many pattern identification applications such as texture analysis, fractal dimension analysis, iris recognition, document analysis, image coding and image representation. To obtain the Gabor filter output the image is convolved with 2D Gabor function given in Eq. 3.7. This function provides the non-orthogonal basis for the wavelet function.

## 3.3 GABOR WAVELET

The 2D Gabor wavelets have been widely used in many image processing applications. Gabor wavelet can be operated as a multi-scale partial differential operator for finding out the edges, corners and blobs in the image. The performance of this type of interest point detector is comparable because of the characteristics of the Gabor wavelets such as rotation invariant and scale invariant. Gabor wavelets are used in estimation of texture features in an image [37-40]. The Gabor function can be used as a non-orthogonal basis sets for construction of wavelets. An image can be expanded using the Gabor wavelet and can efficiently represent the localized frequency description.

A Gabor wavelet is described as a complex oscillating wave modulated by a 2D Gaussian function. Two Gabor wavelets can be differentiated by the ratio of its frequency and the width of the Gaussian function. Gabor wavelet consist of a bank of filters in which each wavelet filter have its discriminating frequency and orientation and when an image is filtered, it provides an estimate of the magnitude of local frequencies present in the image at that parameters of the filter i.e. frequency and orientation.



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

Figure 3.7 2D Gabor wavelet with different value of parameters, (a) $\alpha = 0.5,\ \theta = 0,\ \sigma_x = 1,\ \sigma_y = 1$; (b) $\alpha = 1,\ \theta = 0, \sigma_x = 1,\ \sigma_y = 1$ ; (c) $\alpha = 1,\ \theta = 0, \sigma_x = 0.5, \sigma_y = 2$ ; (d) $\alpha = 1,\ \theta = 45, \sigma_x = 0.5,\ \sigma_y = 2$

For an input image $h(x, y)$, GWT gives a set of images of combination of different scales and orientations and expressed in Eq. 3.9. The function $\psi$ (u,v) represents the mother wavelet.

$$G_{mn}(x, y) = \sum_u \sum_v h(x - u, y - v)\psi^*_{mn}(u, v) \tag{3.9}$$

Where $u, v,$ are the mask sizes of the Gabor filter and $m, n$ represents the scale and orientation, respectively of the mother wavelet function $\psi$, where $\psi^*_{mn}$ is the complex conjugate. The Gabor wavelet filter bank is acquired from the rotation and dilation of the '$\psi$' and is defined as:

$$\psi(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right) e^{\left(\frac{-1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)\right)} e^{2\pi j\omega x} \tag{3.10}$$

The GWT is obtained by the generating function as

$$\psi_{mn}(x, y) = \alpha^{-m}\psi(\overline{x}, \overline{y}) \tag{3.11}$$

Where $m = 0,1,2,3 \ldots \ldots M - 1$ , $n = 0,1,2,3 \ldots \ldots N - 1$ and $\alpha$ is the scale factor of the wavelet function.

$M$: Total number of scales; and

$N$: Total number of orientations

The values of $\overline{x}$ and $\overline{y}$ in the generating function is given as:

$$\overline{x} = \alpha^{-m}(xcos\theta + ysin\theta) \tag{3.12}$$

$$\overline{y} = \alpha^{-m}(-xsin\theta + ycos\theta), \qquad \text{for} \quad \alpha > 1 \text{ and } \theta = \frac{n\pi}{N}.$$

Fig. 3.8 depicts the Gabor wavelet filter for different values of the parameters $\alpha, \theta, \sigma_x \; and \; \sigma_y$.

Gabor wavelet is widely used for extracting features from an image for a recognition system. Two different types of features are extracted from the image, texture based and local key points based. A bank of Gabor wavelet filter is convolved with a resultant image consisting of magnitude and local frequencies of that value of scale and orientation.

Texture based information can be obtained by calculating the energy of the magnitude response of the filtered image. Gabor features are calculated in terms of the mean and variance of the filtered image at various scales and orientations. If we consider the total number of scales as $S$ and orientations as *T*, then the energy content of the filtered image of size $M \times N$ at scale s and orientation r is given in Eq. 3.13.

$$\mathbb{E}(s,t) = \sum_x \sum_y \left| G_{st}(x,y) \right| \tag{3.13}$$

The mean $\mu_{st}$ and the standard deviation $\sigma_{st}$ of all the filtered images are computed using the following equations.

$$\mu_{mn} = \frac{\mathbb{E}(s,t)}{MN} \tag{3.14}$$

$$\sigma_{st} = \sqrt{\frac{\mathbb{E}(s,t) - \mu_{s,t}}{MN}} \tag{3.15}$$

A feature vector ( $FG$) representing a particular image is defined using these statistical parameters of the filtered image as feature components and is given as:

$$FG = (\mu_{00}, \sigma_{00}, \mu_{01}, \sigma_{01}, \ldots \ldots \mu_{M-1\,N-1}, \sigma_{M-1\,N-1}) \tag{3.16}$$

This feature vector provides robustness to the system due to its scale and rotation invariant property.

The local key point based information is based on allocating jets [60] over the filtered image from Gabor wavelet filter bank. A jet is basically considered as an image point information consisting of magnitude as well as phase value as given in Eq. 3.17 where $J_i$ represents the value for the i$^{th}$ wavelet filter.

$$J_i = a_i \, e^{j\emptyset_i} \tag{3.17}$$

The phase information is coded because the variation is phase with location is greater than compared to the magnitude. Also, they are used to discriminate the points where the value of the magnitude is same or nearly equal. These jet points can be further modeled using graph based techniques to obtain the information regarding the relation between the points, hence representing the structure of the object. Fig. 3.8 shows the feature points on the face image used to determine the jet vectors.

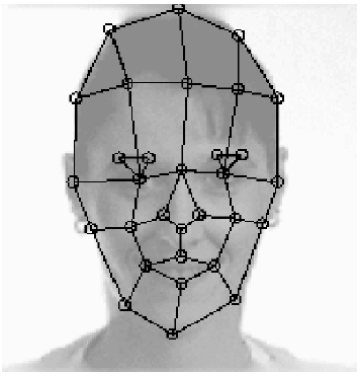


Figure 3.8 The jet points on a human face where the points are modeled by graph based approach as proposed by Laurenz Wiskott et al. in 1997

## 3.4 RADON TRANSFORM

Radon transform is widely used in the field of image processing such as image retrieval, denoising of image, image watermarking, image reconstruction, line detection, image restoration and image hashing. In the most general form the Radon transform describes the shape of the object in an image. Radon transform computes the integral of the projection of the image pixel over a set of lines at a particular direction. The result is in a form of 2D image in terms of angle θ and space. Radon transform is mathematically represented by Eq. 3.18 where the point of integration is given by ρ.

$$R(\rho,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u,v)\,\delta(\rho - u\,\cos\theta - v\,\sin\theta)dxdy \tag{3.18a}$$

Where , $\rho = \mathrm{u}\cos\theta + v\,\sin\theta$ (3.18b)

The Radon transform is a conversion from the Cartesian rectangular coordinates (x,y) to a polar coordinates (ρ,θ) i.e. distance and angle. For the computation of Radon we consider an array of source. The concept of Radon is better understood by Fig. 3.9. Fig 3.9 (a) describes that for every angle θ and at distance ρ, the pixels are integrated along the perpendicular line. The Fig 3.9(b) depicts a straight, thick line in the image; the resulting Radon projection at angle 19 is spread over a large area as shown in Fig. 3.9(c); the Fig 3.9(d) shows the complete Radon transform of the image. The center white point depicts the maximum value of the integral and the distance is from the center of the image, thus it represents the slope of the line in the image along with its position.

For the computation of Radon transform the image is centered by subtracting half f the width and half of the height from each image pixel coordinates. Fig. 3.9 depicts graphically, a line at angle $\theta$ and the corresponding x-min and y-min values for which the image has to be integrated. The equation of the integration line is given as Eq. 3.19

$$y = mx + b \tag{3.19}$$

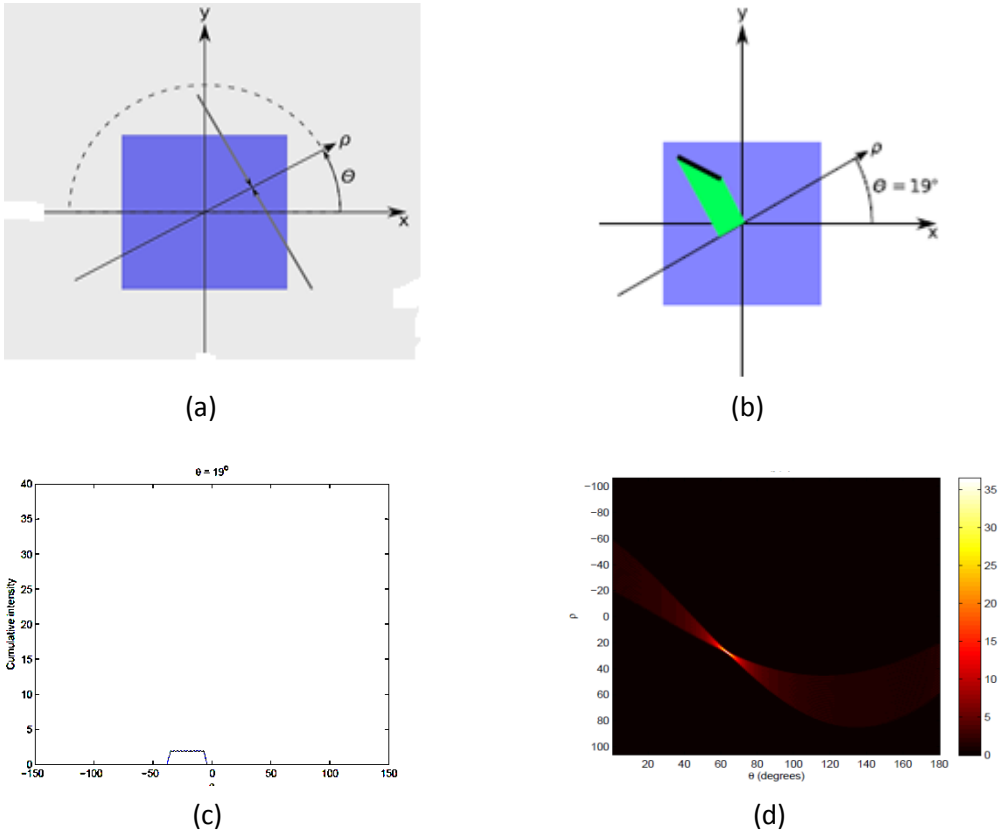



Figure 3.9 Computation of Radon Transform

Where a is the slope of the line and b is the intersection along the y axis given as follow:

$$m = -\frac{\cos\Theta}{\sin\Theta} \quad \text{and} \quad b = \rho \sin\Theta \tag{3.20}$$

The value of m and b are determined for each pair of values of $\Theta$ and p. The maximum value of p is determined by the diagonal value to the image. For reducing the calculation we determine the minimum and maximum value along the y axis. And hence for its

determination the calculation depends on the area in which $\Theta$ lies. The region is divided into 4 sections as shown in Fig. 3.10.

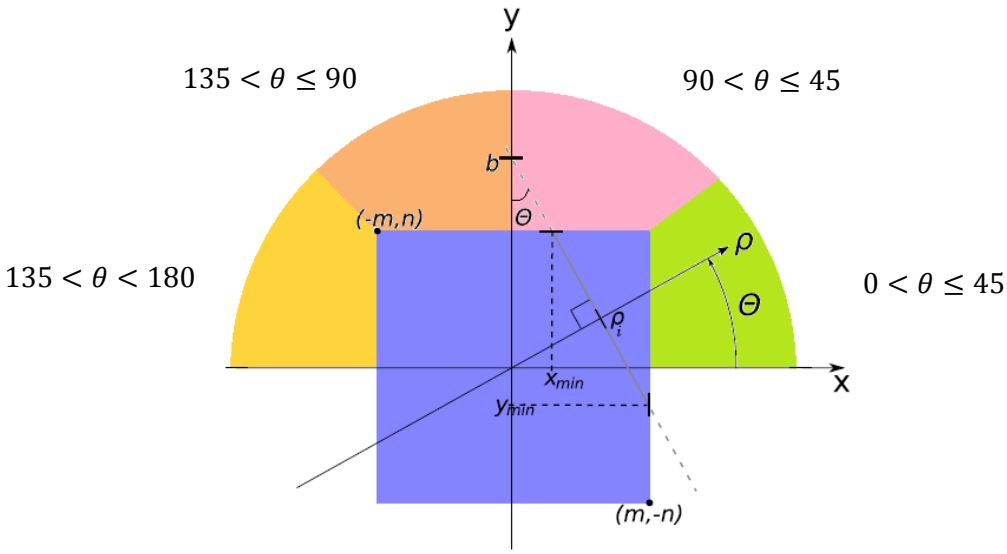



Figure 3.10 Radon computation shown for region $0< \theta \leq 45$

The computation of y minimum and y maximum for different region is given below.

For $0 < \Theta \leq 45$

$$x = \frac{y-b}{a} \tag{3.21a}$$

$$y_{min} = \max(-n, am + b) \tag{3.21b}$$

$$y_{max} = \min(n, -am + b) \tag{3.21c}$$

For $45< \Theta \leq 90$

$$x = ax + b \tag{3.22a}$$

$$y_{min} = \max\left(-m, \frac{n-b}{a}\right) \tag{3.22b}$$

$$y_{max} = \min\left(-m, \frac{-n-b}{a}\right) \tag{3.22c}$$

For $90< \Theta \leq 135$

$$x = ax + b \tag{3.23a}$$

$$y_{min} = \max\left(-m, \frac{-n-b}{a}\right) \quad (3.23b)$$

$$y_{max} = \min\left(m, \frac{-n-b}{a}\right) \quad (3.23c)$$

For $90 < \Theta \leq 135$

$$x = \frac{y-b}{a} \quad (3.24a)$$

$$y_{min} = \max(-n, -am + b) \quad (3.24b)$$

$$y_{max} = \min(n, am + b) \quad (3.24c)$$

For $\Theta = 180$

$$\rho = x + \frac{\rho_{max}-2m}{2} \quad (3.25a)$$

$$y = [-m, m] \quad (3.25b)$$

The value of $\rho_{max}$ is obtained by the image half width and half height as

$$\rho_{max} = \sqrt{(2m)^2 + (2n)^2} \quad (3.25c)$$

As in the equation for x determines that it depends on the chosen y coordinate such that all pixel along the perpendicular direction is taken into consideration when the transform is calculated.

## 3.5 DWT ON RADON TRANSFORM

A wavelet transform is widely used for the analysis of zero dimensional singularity i.e suitable for determining the point like phenomena, but in case for dimensions such as 2 or higher the wavelet transform are poorly modeled. In the case of 2D singularity the Radon transform is employed. The Radon transform converts the point information into line information, i.e. a point singularity converts into line singularity. Applying DWT on Radon transform of images of higher dimension creates great complexities, thus a dimension reduction operation would be required. For making the algorithm simpler, we construct the

concept of discrete Radon transform, i.e. the number of projection angles would be less as compared to the standard no of projection angle equal to $180^{\circ}$. For example, if we consider an image of size $m \times m$, then the number of projections required to represent the Radon transform is equivalent to m+1. Thus the angle variation from $0^{\circ} - 180^{\circ}$ is divided into m+1 equidistant projections. The application of DWT is represented in Fig.3.11.

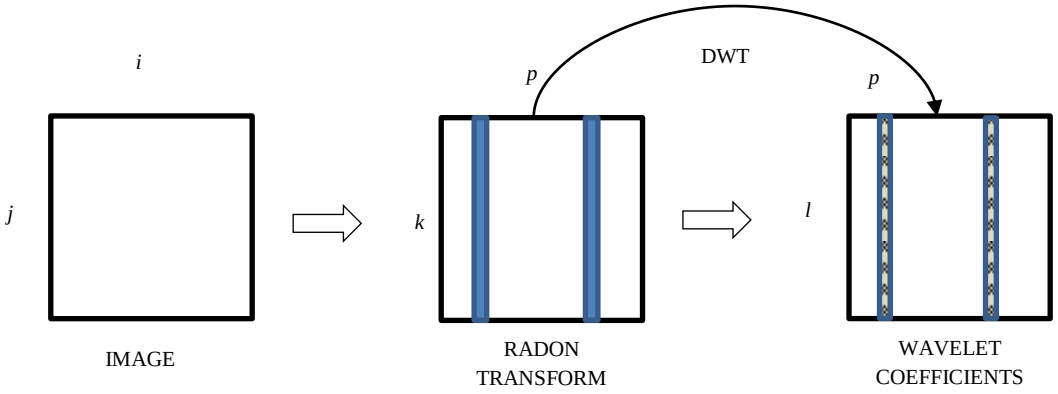



Figure 3.11 Application of DWT on Radon projection

Discrete Radon Transform of an image $h$ of grid $m \times m$ is defined as

$$r_p[k] = RA_h(p,k) = \frac{1}{\sqrt{m}} \sum_{(i,j) \in L_{p,k}} h[i,j] \tag{3.26}$$

Here $L_{p,k}$ represents the set of points that determines the line on the image $h$

$$L_{p,k} = \{(i,j): j = pi + k \ (mod\ m),\ i \epsilon Z_m\}, 0 \leq p < m, \tag{3.27a}$$

$$L_{m,k} = \{(k,j):\ j \epsilon Z_m\} \tag{3.27b}$$

Where $L_{m,k}$ corresponds to the cases of vertical line.

Now the Radon Wavelet coefficient is calculated by applying 1D-DWT on each of the Discrete Radon transform sequence of each direction (p), $r_p[0], r_p[1], r_p[2], r_p[3] \ldots \ldots r_p[m-1]$. This is represented by Fig. 3.11 and expressed as:

$$RW_h(p,l) = DWT(r_p) \tag{3.28}$$

These Radon wavelet coefficients are widely used in extracting the features for the process of image recognition. The multi-resolution analysis provided by wavelet transform gives the frequency related information for that particular resolution. The accurate result can be obtained by considering the scale for which the variation of the magnitude and phase response for different classes is strong enough for classification.

The Radon Wavelet coefficient provides information about the edges in the image at different scale and a frequency is also assigned to it. A 2D image is represented by 2D Radon Wavelet coefficients and for the process of forming the feature vector this dimension has to be reduced into 1D. The most important features that can be extracted from Radon wavelet coefficients is by reshaping the lowest (LL) frequency component obtained by application of single stage DWT [50]. Feature vector can be obtained from this LL part.

Another method to obtain 1D feature vector is by square integrating the Radon Wavelet coefficients along each projection. Thus, for a particular projection $p$ the feature point is given by Eq. 14

$$FR(p) = \sum_l RW^2(p,l) \tag{3.29}$$

If we consider the image dimension as $m \times m$, then the length of feature vector is $m+1$ as the number of projections in FRIT is $m+1$.

## 3.6 FORMATION OF UNIFIED DESCRIPTOR

The features extracted from both the Gabor wavelet transform and Radon Wavelet transform are combined together, giving a unified feature vector that is robust and invariant to scale, rotation and translation of the image column for a single image. The steps to extract the unified descriptor feature vector are described in Fig. 3.12 for an image of Cambridge hand gesture data set.

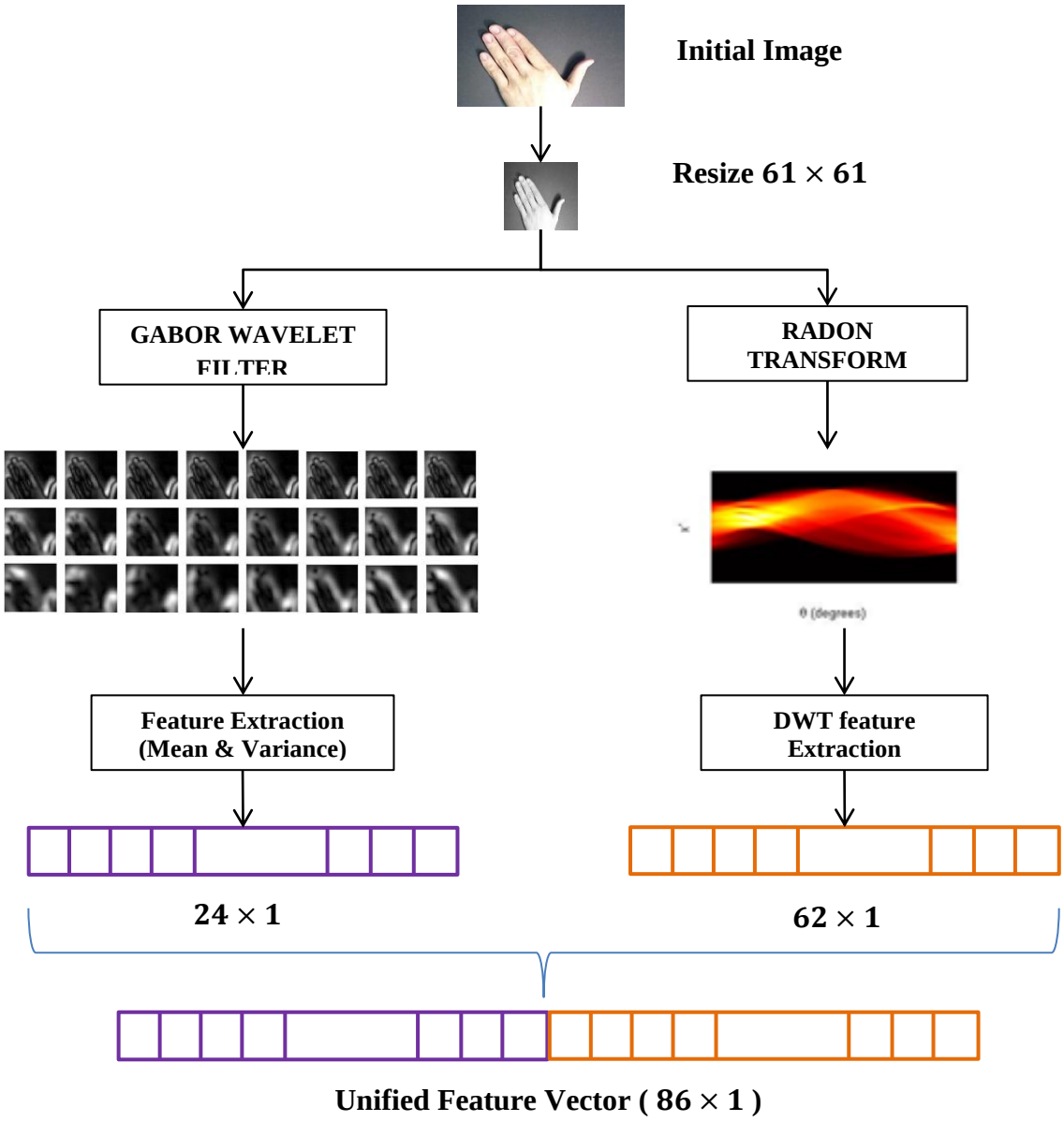



Figure 3.12 Feature extraction flow diagram

The features extracted from the filtered images of Gabor Wavelet is of length $N \times 1$ where $N$ correspond to the total number of the Gabor wavelet filters and usually given by $N = S \times T$, where $S$ is the total number of scale and $T$ is the total number of orientations used for developing Gabor wavelet filter bank. The feature vector obtained from the Radon wavelet transform depends on the size of the image. If the size of the image is $m \times m$ then the length of the feature vector is $m + 1$. The total length of the feature vector is obtained by Eq. 3.16 and Eq. 3.29 and is equal to

$$L = (N + m + 1) \times 1 \tag{3.30}$$

The descriptor in tested on varied data sets for applications such as hand gesture and human activity recognition.

## 3.7 CLASSIFIER

Classification refers to the process of categorizing data by analyzing the features which distinguish one class from the other classes. A classifier is developed on the bases of the learning approach adopted by the classifier. There are two types of learning: supervised learning and unsupervised learning. In supervised learning the system already has the prior knowledge about the class of the training data. This information develops the classifier based on the training data and then the developed classifier is used to determine the class of the test data. To develop a system a minimum amount of training data is required to train the classifier to give optimal performance, even over-training the classifier also leads to the incorrect classification of data. In unsupervised learning the prior knowledge about the class of the data is not available and the data are classified on the basis of clustering the features into different classes. The major application of unsupervised learning is in the field of segmentation of image and speech coding. The standard classifier being used are SVM, KNN decision tree and Naïve Bayes. In our work we have used supervised learning technique as the classifier have the prior knowledge about the class label of the training data and the classifier used are SVM and KNN to determine the class of the test data.

### 3.7.1 Support Vector Machine

SVM is a supervised learning technique and the label of the training data is used to obtain a decision boundary between the two classes. It is a non-probabilistic binary linear classifier and develops a hyperplane in higher dimensional space which distinguish between the classes. The hyperplane is developed by maximizing the margin between the two classes. A margin is the distance between the two classes and is determined by the support vector points. The support vectors are the points that are nearest to the hyperplane. Maximizing the margin leads to the linear programming problem. This classifier is also known as a maximum margin classifier.

- **Separable classes**

In this we consider a two class problem which is separable. Let $X$ be training set consisting of $N$ feature vectors, i.e. $X = [x_1, x_2, x_3, \dots \dots \dots x_N]$. This training set consists of data from 2 classes $w_1$ and $w_2$, and which is assumed to be linearly separable. The equation of the hyperplane that correctly classify all the training data set is given by Eq. 3.31.

$$g(x) = w^T x + w_o = 0 \tag{3.31}$$

Fig. 3.13 explains the classification task with different possible linear hyperplanes. Now we define the support vector as the points nearest to the hyperplane. Two planes are constructed one for the each class that passes through the support vector points of the respective class. Then the hyperplane for classification is developed which is equidistant from the initial two planes obtained from the support vector of the two classes. The algorithm is developed that maximizes the distance from the support vectors of both the classes. This leads to the construction of a linear programming problem that consists of the objective to maximize the margin with the constrain that the trained data must be correctly classified by the obtained hyperplane.

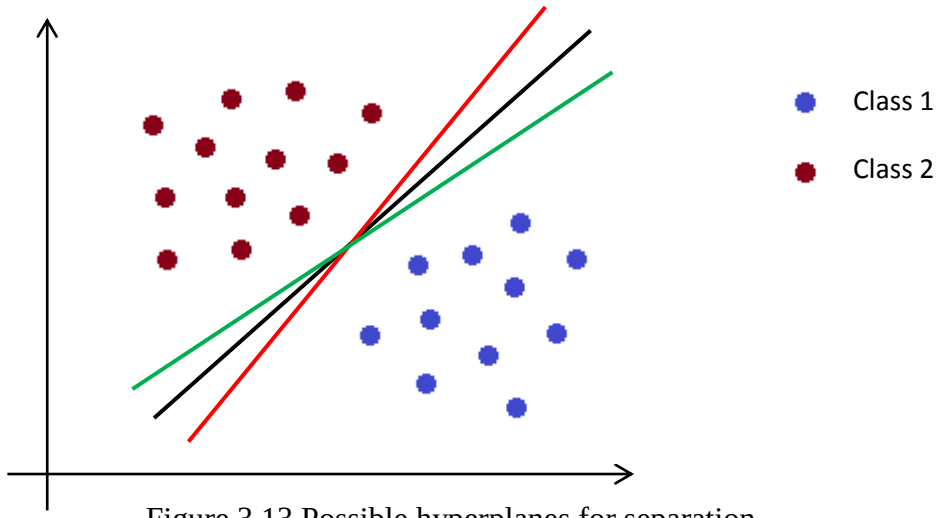



Figure 3.13 Possible hyperplanes for separation

Now consider the Fig. 3.14, in which support vectors are encircled and a line is passed through it. The figure depicts the two possible directions of the hyperplane with margins $2z_1$ and $2z_2$. The goal is to select the hyperplane that maximizes the margin and from the figure it can be deduced that direction 2 have larger margin than direction 1, thus direction 2 hyperplane is considered to give optimal performance. The distance of a point from the hyperplane is given by Eq. 3.32.

$$z = \frac{|g(x)|}{\|w\|} \tag{3.32}$$

The values of $\boldsymbol{w}$, $w_0$ is scaled such that value of $g(x)$ is equal to 1 for support vectors in $w_1$ and -1 for $w_2$. The mathematical formulation is given by a linear programming problem as follows:

Having margin as : $$\frac{1}{\|w\|} + \frac{1}{\|w\|} = \frac{2}{\|w\|} \tag{3.33}$$

Requiring that : $$w^T x + w_o \geq 1\,, \quad \forall\, x \in w_1 \tag{3.34a}$$

$$w^T x + w_o \leq -1\,, \quad \forall\, x \in w_2 \tag{3.34b}$$

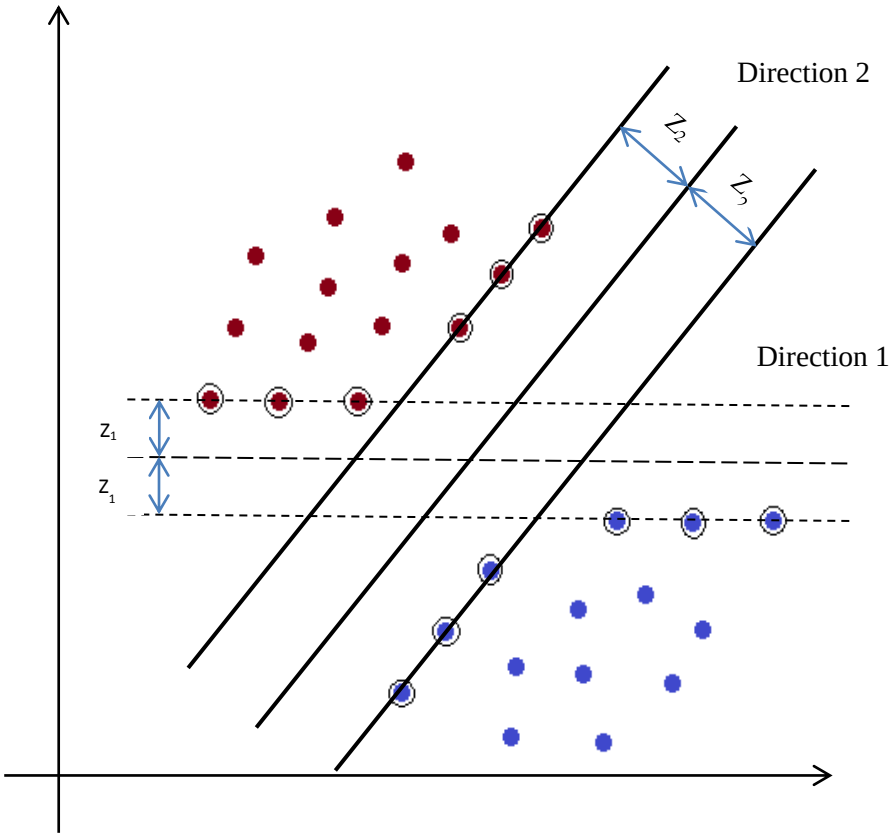



Figure 3.14 Two possible hyperplane for a support vector system

Now consider $Y = [y_1, y_2, y_3, \ldots\ldots\ldots y_n]$ as the class indicator for $X$ whose value is +1 for $w_1$ and -1 for $w_2$. Compute the parameter $w, w_0$ of the hyperplane so that to

$$\text{Minimize} \qquad J(w, w_0) = \frac{1}{2}\|w\|^2 \tag{3.35}$$

$$\text{Subject to} \qquad y_i(w^T x_i + w_0) \geq 1, \quad i = 1,2 \ldots, N \tag{3.36}$$

The above equations define a non linear quadractic optimization subject to linear inequality constraints. This can be solved by Karush-Kuhn-Tucker (KKT) conditions and the result is obtained in terms of Lagrange multiplier $\alpha_i$ and Lagrangian function $L(w, w_0, \alpha)$ as:

$$L(w, w_0, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{N} \alpha_i [y_i (w^T x_i + w_0) - 1] \quad (3.37)$$

Applying KKT conditions gives:

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i \quad (3.38)$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \quad (3.39)$$

$$\alpha_i \geq 0 \quad (3.40)$$

Substituting Eq. 3.38 and 3.39 into 3.37 results in an optimized equation for the classifier as given in Eq. 3.41

$$\max_{\alpha} \left( \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{ij} \alpha_i \alpha_j y_i y_j x_i^T x_j \right) \quad (3.41)$$

$$\text{subject to} \quad \sum_{i=1}^{N} \alpha_i y_i = 0 \quad (3.42)$$

$$\alpha_i \geq 0 \quad (3.43)$$

After finding out Lagrange's coefficient values from Eq. 3.42, one can find out the hyperplane from Eq. 3.38.

- **Nonseparable classes**

One special case lies in the case of SVM where the classes are not separable as given in Fig. 3.15. The margin is defined by the distance between the parallel hyperplanes described by Eq 3.35 and 3.36. Thus Eq. 3.44 is obtained.

$$w^T x + w_o = \pm 1 \quad (3.44)$$

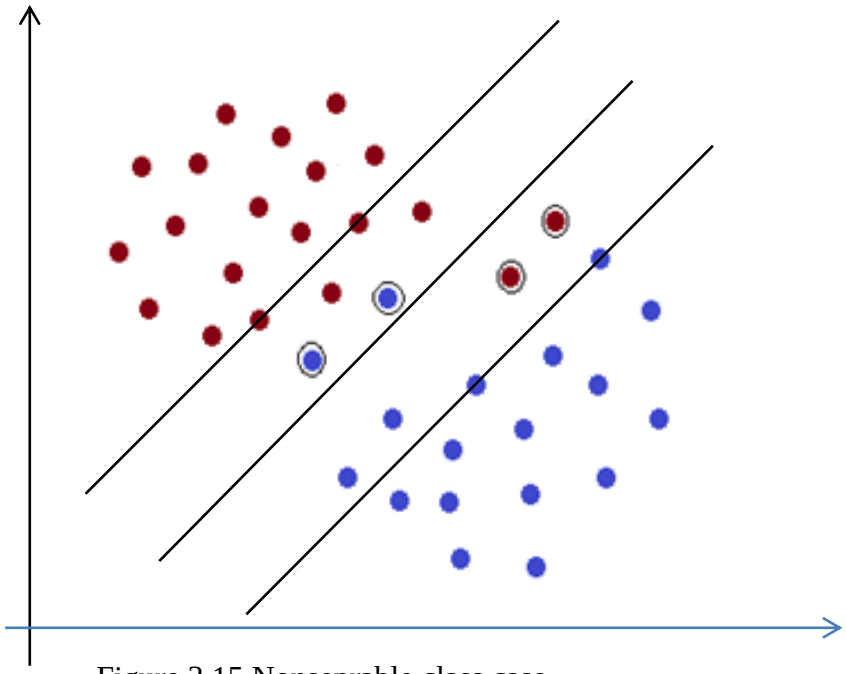


Figure 3.15 Nonseprable class case

The training feature vectors are now categorized into 3 cases:

- Feature vectors that fall outside the band and are correctly classified and satisfies the constraint in Eq. 3.36.
- Feature vectors lying inside the two parallel lines and classified correctly. They satisfy the inequality

$$0 \leq y_i(w^T x + w_0) < 1 \tag{3.45}$$

- Feature vectors that are misclassified and satisfy the inequality

$$y_i(w^T x + w_0) < 0 \tag{3.46}$$

The above cases can be resolved by a single constraint by introducing a new set of variable $\xi_i$ known as slack variable.

$$y_i(w^T x + w_0) < 1 - \xi_i \tag{3.47}$$

In the first case $\xi_i = 0$, second case $0 < \xi_i \leq 1$, and the third $\xi_i > 1$.

The equation derived for solving this condition is given by

$$J(w, w_0) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} I(\xi_i) \tag{3.48}$$

Where $I(\xi_i)$ is given by

$$I(\xi_i)\begin{cases}1 & \xi_i > 0\\ 0 & \xi_i = 0\end{cases} \tag{3.49}$$

Where C is a positive constant that influence the two competing terms. The optimization equation is obtained as follows:

$$\max_{\alpha}\left(\sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{ij}\alpha_i\alpha_j y_i y_j x_i^T x_j\right) \tag{3.50}$$

$$\text{Subject to} \qquad 0 \le \alpha_i \le C,\ i = 1,2,\dots.,N \tag{3.51}$$

$$\sum_{i=1}^{N}\alpha_i y_i = 0 \tag{3.52}$$

In this the Lagrange multipliers correspond to the vector points between the two margins i.e either within the margin or on the wrong side of the hyperplane.

- **Multiclass case**

In this we consider the case of M different classes for developing the classifier. One method is to create a set of M- two class problem known as one against all. For every class a discriminant function is designed $g_i(x), i = 1,2,\dots M$, so that $g_i(x) > g_j(x), \forall_j \neq i, if\ x \epsilon w_i$. The optimal discriminant functions are designed such that $g_i = 0$ is the optimal hyperplane seprating class $w_i$ from all other classes. Thus each classifier satify

$$g_i(x) > 0\ for \quad x \epsilon w_i \quad \text{and} \tag{3.53a}$$

$$g_i(x) < 0 \ \text{ otherwise} \tag{3.53b}$$

The classification is obtained by the following rule:

$$assign\ x\ in\ w_i\ if\ i = \arg\max_k [g_k(x)]$$

The drawback of this technique is that it can lead to more that one positive $g_i(x)$. The problem is asymmetric as the number of negative classes is much greater than positive classes.

Another alternative method is the one-against-one. In this $M(M-1)/2$ binary classifiers are trained and separates a pair of classes. The decision is made according to the majority vote of the classification result. The disadvantage is that a large no of classifiers has to be developed.

### 3.7.2 K Nearest Neighbours

The classifier is based on the similarity measurement between the test sample and its K nearest neighbour. It is a nonlinear classifier and follows the nearest neighbour rule. It is a non-parametric method as there is not learning stage of the classifier. The information of the training and test data is given to the classifier at the same instance and the classifier determines the class of the test data.

Many distance functions are available to measure the similarity and its usage is dependent on the type of the features. Euclidean distance, Mahalanobis distance, Hamming distance are some of the distance functions. The type of distance measurement also leads to determine the performance of the classifier. The KNN algorithm gives the simplest machine learning classifier also known as instance based learning classifier.

Consider an example of two classes given in Fig.3.16. The unknown sample is given by green colour. If we consider K=1, i.e. the class of the unknown test data is determined as diamond as it is nearer and if K=3, the class is determined as square as there are 2 squares nearer to the test data.

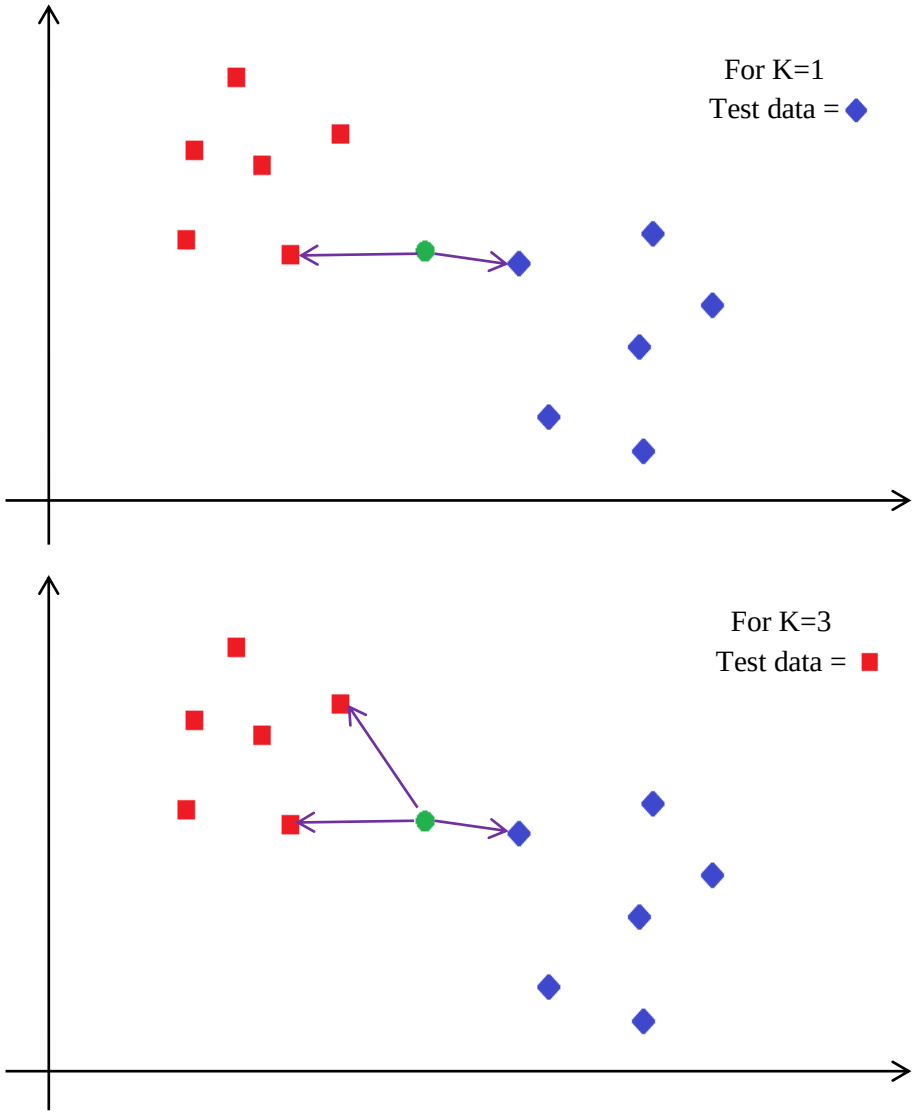



Figure 3.16 KNN classifier

The steps for classification includes: choosing the value of K and the distance measurement, finding the distance between the test sample and the training data and hence determine the nearest points and finally finding the majority class among the K-nearest neighbour. The steps are summarized below.

- K nearest training neighbors are identified, regardless of the class label. The value of K is chosen as odd for a two class problem and in general not a multiple of the number of classes M.
- Out of these K samples, the number $k_i$ of vectors belonging to class $w_i, i = 1,2,, \ldots M$.
- The test vector x is assigned to the class $w_i$ With a maximum value $k_i$.

# CHAPTER 4
# EXPERIMENT AND RESULTS

The performance of the unified feature descriptor obtained in this work using Gabor wavelet and Radon wavelet transform is obtained by testing on hand gesture and human activity recognition data sets. The hand gesture data sets tested in this work are Jochen Triesch, Cambridge and NUS. The human activity dataset tested are KTH, Weizmann and Ballet. The classifier used is KNN and SVM for classification. The ARR is calculated to check the performance of the descriptor and is compared with the present state of the art. The details of the data sets, the training and testing sets and the results obtained are further described in this chapter.

The Jochen Triech hand posture dataset consist of 10 different class of images i.e a, b, c, d, g, h, i, l, v, y and of 24 subjects, under three types of background i.e. light, dark and complex. The light background images are shown in Fig.4.1. The classifier is trained by 60 images and 12 images are used for testing, i.e. 20 images of same background are trained and 4 images are tested. The classifier used here is SVM and the confusion matrix is plotted in Table 4.1. The ARR obtained is 97.5%.

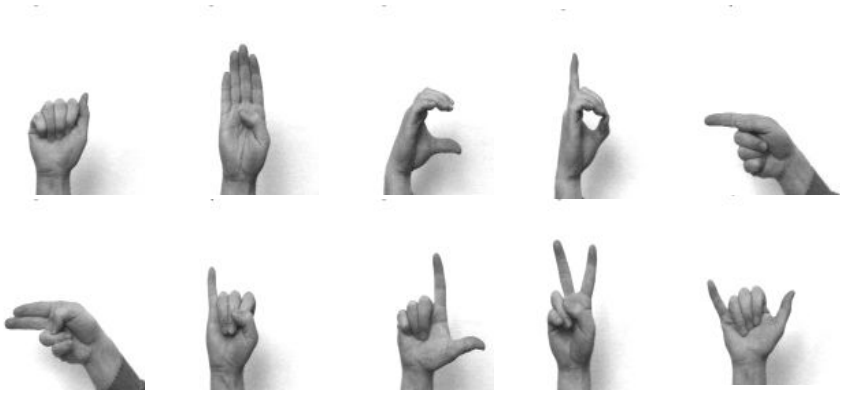


Figure 4.1 Jochen Triech data set

| Predicted class / Actual class | a | b | c | d | g | h | i | l | v | y |
|---|---|---|---|---|---|---|---|---|---|---|
| **a** | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **b** | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **c** | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **d** | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| **g** | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| **h** | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 |
| **i** | 0 | 0 | 0 | 12.5 | 0 | 0 | 87.5 | 0 | 0 | 0 |
| **l** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 87.5 | 0 | 12.5 |
| **v** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 |
| **y** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |

Table 4.1 Confusion matrix for Jochen Triech data set using SVM classiifer

| Approach | Features Used | ARR (%) |
|---|---|---|
| Jochen Triech et al [61] | Graph Matching | 93.8 |
| Pramod Kumar et. al. [62] | Visual cortex | 96.1 |
| *Our Method* | *Gabor-Radon Transform* | *97.5* |

Table 4.2 Comparison of ARR with the techniques of others for Jochen Triech data set

The NUS hand posture data set consist of 10 different postures with each posture consisting of images from 24 sample images. The images are captured by varying the position and size of the hand within the image frame. The images of the data set are shown in Fig. 4.2. The SVM classifier is used for evaluating the ARR, in which 20 images are used

for training and 4 images are used for testing. The process is repeated 3 times with randomly varying the training and testing images. The confusion matrix is shown in Table 4.3. The ARR obtained is 92.5%.

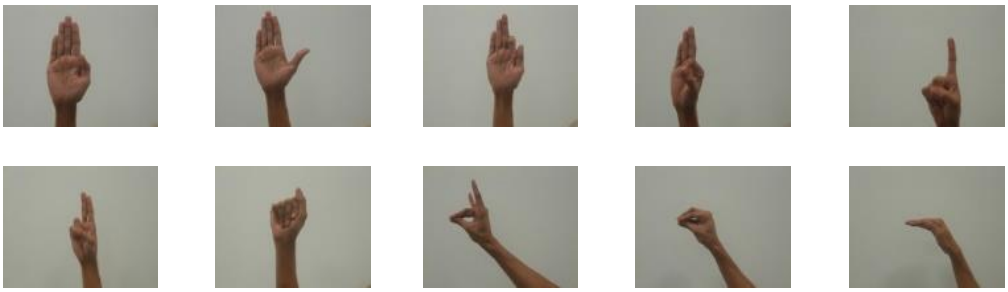


Figure 4.2 NUS Hand gesture data set

| Predicted class / Actual class | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **2** | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **3** | 0 | 8.33 | 91.66 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **4** | 16.66 | 0 | 33.33 | 50 | 0 | 0 | 0 | 0 | 0 | 0 |
| **5** | 0 | 0 | 0 | 0 | 91.66 | 0 | 8.33 | 0 | 0 | 0 |
| **6** | 0 | 0 | 8.33 | 0 | 0 | 91.66 | 0 | 0 | 0 | 0 |
| **7** | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 |
| **8** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| **9** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 |
| **10** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |

Table 4.3 Confusion Matrix for NUS hand gesture data set using SVM classifier

| Approach | Features Used | ARR (%) |
|---|---|---|
| Pramod Kumar et. al. [62] | Visual cortex | 92.15 |
| *Our Method* | *Gabor-Radon Transform* | *92.5* |

Table 4.4 Comparison of ARR with the techniques of others for NUS data set

The Cambridge hand gesture data set consist of 9 action classes and a total of 900 image sequence, i.e. 100 image sequence for each activity class. The data set consist of 3 primitive shapes: flat, spread and V-shape; and 3 primitive motions: leftward, rightward and contract as given in Fig. 4.3. From the Fig. 4.3, similarity along the row can be easily observed. It consists of 5 different illumination conditions in which 2 subjects perform the action in 10 arbitrary motions.

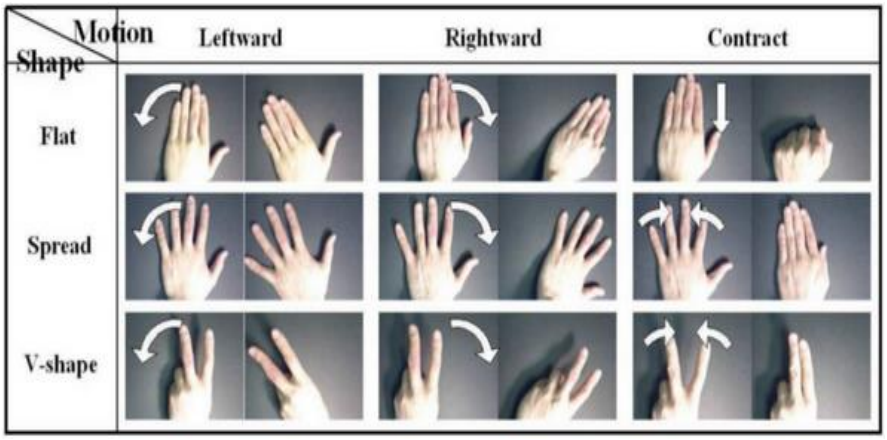



Figure 4.3 Cambridge data set

To test the unified descriptor proposed in our work, one frame near to the final state of the motion is selected from each of the 100 image sequence for a particular class and similarly for other classes. The frame near to the final state is considered as from the Fig. 4.3, the similarity of the initial image sequence along the row can be easily observed. Now we have a total of 900 images, i.e 100 images for each action class for feature extraction and classification process. The unified feature is extracted and then used for classification by an SVM classifier using 80 images for training from 4 out of 5 sets and 20 images for testing from the $5^{th}$ set out of the 100 images of a class of action. The confusion matrix is obtained in Table 4.5 using SVM classifier. The ARR obtained is 91.11% and compared with the other state of art in Table 4.6.

| Predicted class / Actual class | Flat Leftward | Flat Rightward | Flat Contract | Spread Leftward | Spread Rightward | Spread Contract | V-shape Leftward | V-shape Rightward | V-shape Contract |
|---|---|---|---|---|---|---|---|---|---|
| **Flat Leftward** | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Flat Rightward** | 0 | 90 | 0 | 0 | 5 | 5 | 0 | 0 | 0 |
| **Flat Contract** | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Spread Leftward** | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| **Spread Rightward** | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 |
| **Spread Contract** | 5 | 0 | 10 | 0 | 5 | 80 | 0 | 0 | 0 |
| **V-shape Leftward** | 0 | 0 | 0 | 10 | 0 | 0 | 90 | 0 | 0 |
| **V-shape Rightward** | 0 | 0 | 10 | 0 | 10 | 0 | 0 | 80 | 0 |
| **V-shape Contract** | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 80 |

Table 4.5 Confusion matrix for Cambridge dataset using SVM classifier

| Approach | Features Used | ARR (%) |
|---|---|---|
| Kim et al [63] | Correlation Analysis | 82+ 3.5 |
| Harandi et al. [64] | RLPP | 86.3 + 1.3 |
| Lui et al. [65] | Product Manifolds | 88 |
| *Our Method* | *Gabor-Radon Transform* | *91.11* |

Table 4.6 Comparison of ARR with the techniques of others for Cambridge data set

Now, the human activity data sets are experimented to determine the accuracy of the descriptor. Aggrawal and Ryoo [66], explained the different approaches being used for analyzing human activity in a video. KTH dataset given by Schuldt et al. [67] which consist of 6 classes of human activities, i.e. running, jogging, walking, hand clapping, hand waving

and boxing. Each action class is performed by 25 persons in four different scenarios. The video sequences for different activity is shown in Fig. 4.4.

In this 20 key frames are deduced from each video. The training image set is obtained by deducing key frames from the videos of training set. The confusion matrix is obtained in Table 4.7. The ARR come out to be 96.66%, which is much higher than the previously received recognition by different techniques. The confusion between walking, running and jogging is due to the similarity in the poses for these activities. The results are compared with the previous existing techniques in Table 4.8 and shows that our method is superior with respect to the other techniques.

Figure 4.4 KTH data set

| Predicted Class / Actual Class | Jogging | Walking | Running | Hand Clapping | Hand Waving | Hand Boxing |
|---|---|---|---|---|---|---|
| **Jogging** | 95 | 5 | 0 | 0 | 0 | 0 |
| **Walking** | 5 | 95 | 0 | 0 | 0 | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Running** | 5 | 5 | 90 | 0 | 0 | 0 |
| **Hand Clapping** | 0 | 0 | 0 | 100 | 0 | 0 |
| **Hand Waving** | 0 | 0 | 0 | 0 | 100 | 0 |
| **Hand Boxing** | 0 | 0 | 0 | 0 | 0 | 100 |

Table 4.7 Confusion matrix for KTH Dataset using KNN classifier

| Approach | Features Used | ARR (%) |
|---|---|---|
| S.Y. Cho et al [68] | HOG & HOF | 94.55 |
| Mohsen Soryani et al [69] | Body Posture Graph | 94.60 |
| Biyun Sheng et al [70] | HOG Feature Directional Pairs | 94.99 |
| *Our Method* | *Gabor-Radon Transform* | *96.66* |

Table 4.8 Comparison of ARR with the techniques of others for KTH data set

The Weizmann dataset consist of 10 human actions, i.e. walking, running, jumping, galloping sideways, bending, one hand waving, two hands waving, jumping in place, jumping, jack and skipping. These data actions are performed by 9 actors thus there are 9 videos for a single action. The training is done by obtaining key frames from 8 videos and tested on key frames from the single video. The confusion matrix is shown in Table 4.9. The ARR obtained is 96% and compared with the other state of art in Table 4.10. The recognition rate in skip is less because of its similarity with the other actions like walking running and jumping.

Figure 4.5 Weizmann data set

| Predicted class / Actual class | Bend | Jack | Jump | P. Jump | Run | Side | Skip | Walk | Wave 1 | Wave 2 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Bend** | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Jack** | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Jump** | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **P. Jump** | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Run** | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| **Side** | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 |
| **Skip** | 0 | 0 | 20 | 0 | 10 | 0 | 60 | 10 | 0 | 0 |
| **Walk** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| **Wave 1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 |
| **Wave 2** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |

Table 4.9 Confusion matrix for Weizmann data set using KNN classifier

| Approach | Features Used | ARR (%) |
|---|---|---|
| J. Arunnehru et al [71] | Frame Differencing Motion Feature | 92 |
| Li Chuanzhen et al [72] | Spatio Temporal | 92.53 |
| Z. Zhang et al [73] | Motion Features | 92.8 |
| *Our method* | *Gabor-Radon Transform* | *96* |

Table 4.10 Comparison of ARR with the techniques of others for Weizmann data set

# CHAPTER 5
# CONCLUSION AND FUTURE SCOPE

In this thesis a unified global descriptor based on Gabor Wavelet and Radon-Wavelet is proposed for the purpose of pattern identification. The descriptor is robust as it consists of characteristics such as invariant to scale, translation and rotation. The two types of descriptors, i.e. global and local descriptors are explained in this work. The classifiers used for classifying the feature vectors obtained by local descriptors are complex, such as feature matching, Hausdorff average and maximum likelihood classifier. Therefore, we employed the approach of a global descriptor that uses standard classifier such as SVM and KNN techniques.

The preprocessing techniques like segmentation, edge or contour detection, etc. make the system application specific as well as increase the complexity of the system. Many approaches are developed that require a preprocessing stage. The unified feature descriptor proposed here does not require any preprocessing steps, thus making the system less complex. The use of DWT leads to multi resolution analysis provides a complete frequency description of the image. The descriptor has been tested on hand gesture and human activity data set and provides better ARR as compared to the other state of art. The performance of the descriptor can be verified for many other applications in the field of pattern identification.

First order statistical features are obtained from the filtered images of Gabor wavelet transform to define the descriptor, one can work on second order statistic for such as correlation matrix as a future scope. Local feature point extraction can also be studied further. Similarly, in case of Radon-wavelet transform one can obtain features from multi stage decomposition of the Radon transform image.

# REFERENCES


1. Bay, H; Tuytelaars, T.; Gool, V. SURF: Speeded Up Robust Features. Proceedings of Ninth European Conference on Computer Vision, 2006.

2. Donghoon, K.; Trinity, C.; Dublin; Dahyat, R. Face Components Detection using SURF Descriptors and SVMs. International Conference on Machine Vision and Image Processing, 2008; pp 51-60.

3. Lowe, D. G. Distinctive image Features from Scale-Invariant Keypoints. International Journal Computer Vision 2004, 60 (2), 91-110.

4. Lowe, D. G. Object Recognitiom from Local Scale-Invariant features. 7th IEEE International Conference on Computer Vision, 1999; pp 1150-1157.

5. Scovanner, P.; Ali, S.; Shah, M. A 3-Dimensional SIFT Descriptor and its Application to Action Recognition. 15th International Conference on Multimedia,ACM 2007, 357-360.

6. Fernandez, C.; Vicente, M. A. Face recognition using multiple interest point detectors and SIFT descriptors. IEEE international confernce on Automatic face and gesture recognition, 2008; pp 1-7.

7. Dalal, N.; Triggs, B. Histogram of oriented gradients for human detection. IEEE Conference on Computer Vision and Pattern Recognition, 2005; pp 886-893.

8. Tuermer, S.; Kurz, F.; Reinartz, P.; Stilla, U. Airbone vehicle detection in dense urban area using HOG features and disparity maps. 2013, 6 (6), 2327-2337.

9. Gunasekar, S.; Ghosh, J.; Bovik, A. C. Detection on distorted images augmented by perceptual quality-aware features. 2014, 9 (12), 2119-2131.

10. Laptev, I.; Lindeberg, I. Space-time interest points. International conference on Computer Vision.

11. Chenguang, G.; Xianglong, L.; Linfeng, Z.; Xiang, L. A fast and accurate corner detector based Harris algorithm. 3rd international symposium on intelligent information technology application, 2009; pp 49-52.

12. Dai, J.; Pei, L.; Song, W.; Zang, J. Remote sensing image matching via harris detector and SIFT descriptor. 3rd Internatinal confress on Image and Signal Processing, 2010; pp

1-4.

13. Ryu, J. B.; Park, H. H.; Park, J. Corner classification using Harris algorithm. 2011, 47 (9).

14. Dollar, P.; Rabaud, V.; Cottrell, G.; Belongie, S. Behaviour recognition via sparse spatio-temporal features. VS-PETS, 2005.

15. Willems, G.; Tuytelaars, T.; Gool, L. V. An efficient dense and scale invariant spatio-temporal interest point detector. ECCV, 2008.

16. Beaudet, P. Rotation invariant image operators. ICPR, 1978.

17. Lindeberg, T. Feature detection with automatic sclae selection. 1998, 30 (2), 79-116.

18. Bobick, A. F.; Davis, J. W. The recognition of human movements using temporal templates. 2001, 23, 257-267.

19. Sheik, Y.; Sheik, M.; Shah, M. Exploring the space of a human action. IEEE International conference on computer vision, 2005; pp 144-149.

20. Liu, X.; Liu, Q.; Oe, S. Recognizing Non-rigid Human Actions using Joints tracking in space-time. IEEE International confrence on Information Technology:Coding and Computing, 2004; pp 620-624.

21. Lucas, B. D.; Kanade, T. An iterative Image Registration Technique with an application to stereo vision. 7th Internatinal joint conference on artificial intelligence, 1981; pp 24-28.

22. Shi, J.; Tomasi, C. Good Features to Track. IEEE conference on computer vision and pattern recognition, 1994; pp 593-600.

23. Swain, M. J.; Ballard, D. H. Color indexing. International Journal of Computer Vision, 1001.

24. Stricker; Orengo, M. Similarity of colour Images. Storage and Retrieval for image and video databases, 1995.

25. Rui, Y.; Huang, T. S. Image retrieval:Current techniques, Promising Directions and open issues. 1999, 10, 39-62.

26. Liu, F.; Xiong, X.; Chan, L. Image analysis and interpretation. IEEE southwest

Sumposium, 2000.

27. Weng, L.; Preneel, B. Shape-Based features for Image Hashing. IEEE International conference on Multimedia and Expo, 2009.

28. Lisin, D. A.; Mattar, A. M.; Blaschko, B. M.; Benfield, M. C. Combining local and global image features for object class recognition. IEEE conference on computer vision and pattern recognition, 2005.

29. Rahman, S. A.; Cho, S. Y.; Leung, M. K. H. Recognizing human actions by analysing negative spaces. 2011.

30. Rahman, S. A.; Li, L.; Leung, m. K. H. Human action recognition by negative space analysis. Internatinal conference on cyberworld, 2010.

31. Singh, M.; Basu, A.; Mandal, M. K. Human activity recognition based on silhouette directionality. 2008, 18 (9).

32. Myagmarbayar, N.; Yuki, Y.; Imamoglu, N.; Gonzalez, J.; Otake, M.; Yu, w. Human body contour data based activity recognition. 2013.

33. Venkatesha, S.; Turk, M. Human activity recognition using local shape descriptors. 2010.

34. Kumari, S.; Mitra, S. K. Human action recognition using DFT. IEEE national conference on computer vision, Pattern Recognition, Image processing and graphics, 2011; pp 239-242.

35. Pittner, S.; Kamarthi, S. V. Feature Extraction From Wavelet Coefficients for Pattern Recognition Tasks. 1999, 21 (1).

36. Daugman, J. Two-dimensional spectral analysis of cortical receptive field profiles., 2010; pp 847–856.

37. Bovik, A. C.; Gopal, N.; Emmoth, T.; Restrepo, A. Localized measurement of emergent image frequencies by Gabor wavelets. 1992, 38 (1), 691–712.

38. Lee, T. S. Image Representation using 2D Gabor wavelets. 1996, 18 (10).

39. Jiang, W.; Shen, T. Z.; Zhang, J.; Hu, Y.; Wang, X. Y. Gabor Wavelets for Image Processing.; p 110—114.

40. Arivazhagan, S.; Ganesh, L.; Priyal, S. P. Texture classification using Gabor wavelets based rotation invariant features. 2006, 27, 1976-1982.

41. Jain, A. K.; Farrokhina, F. Unsupervised texture segmentation using Gabor filters. 1991, 24 (12), 1167-1186.

42. Dunn, D. F.; Higgins, W. E. Optimal Gabor filters for texture segmentation. 1995, 4 (7), 947–964.

43. Teuner, A.; Pichler, O.; Hosticka, B. J. Unsupervised texture segmentation of images using tuned matched Gabor filters. 1995, 4 (6), 863–870.

44. Pichler, O.; Teuner, A.; Hosticka, B. J. A comparison of texture feature-extraction using adaptive Gabor filtering, pyramidal and treestructured wavelet transforms. 1996, 29 (5), 733–742.

45. Idrissa, O.; Acheroy, M. Texture classification using Gabor filters. 2002, 23 (9), 1095-1102.

46. Ibrahim, M. A.; Hussein, A. W. F.; Mashali, S. A. A blind image restoration system using higher-order statistics and Radon transform. 1998, 523-530.

47. Zhiyong, A.; Zhiyong, Z.; Lihua, Z. Contenet based image retrieval based on the wavelet transform and Radon transform., 2007; pp 1878-1881.

48. Khan, Z. A.; Sohn, W. Hierarchical human activity recognition system bades on R-transform and non-linear kernel discriminant features. 2012, 48 (18).

49. Arivazhagan, S.; Ganesan, L.; Kumar, T.; G Subash. Texture classification using ridgelet transform. 2006, 27, 1875–1883.

50. Kautkar, S.; Koche, ; Keskar, T.; Pande, A.; Rane, M.; Atkinson, G. A. Face Recognition Based on Ridgelet Transforms. CEBT, 2010.

51. Zang, D.; Wang, W.; Gao, W.; Jiang, S. An Effective Local Invariant Descriptor Combining Luminance And Colout. IEEE International Conference On Multimedia And Expo, 2007; pp 1507-1510.

52. Azad, P.; Asfour, T.; Dillmann, R. Combining Harris Interest Points and the SIFT Descriptor for Fast Scale-Invariant Object Recognition. IEEE International Conference on Intelligent Robots and Systems, 2009.

53. Tian, Y. L.; Cao, L.; Liu, Z.; zang, Z. Hierarchical Filtered Motion for Action Recognition in Crowded Videos. 2012, 42 (3).

54. Daubechies, I. Orthogonal bases of compactly supported wavelets. Comm. Pure Appl. Math., vol., 1988.

55. Guo, B.; Jiang, J. A modified shape descriptor in wavlets compressed domain. International conference on Image Processing, 2002; pp 936-939.

56. Mallat, S.; Hwang, W. L. Singularity detection and processing with wavelets. 1992, 38 (2), 617-643.

57. Tieng, Q. M.; Boles, W. W. Wavlet-based affine invariant representation: a tool for recognizing planar objects in 3D space. 1997, No. 8, 846-857.

58. You, X.; Tang, Y. Y. Wavelet-based Approach to character Skeleton. 1220-1231.

59. Prabhakar, C. J.; Kumar, P. U. P. Underwater image denoising using adaptive wavelet subband thresholding., 2010; pp 322-327.

60. Wiskott, L.; Fellous, J. M.; Kruger, N.; Von, C.; Malsburg, D. Face Recognition By Elastic Bunch Graph Matching. 1997, 19 (7), 775-779.

61. Triesch, ; Malsburg, v. d. Robust Classification of Hand Postures against Complex Backgrounds. Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 1996; pp 170-175.

62. P, P. k.; Vadakkepat, P.; Poh, L. A. hand posture and face recognition using a fuzzy rough approach. International Journal of Humanoid Robotics 2010, 331-356.

63. Kim, T. K.; Cipolla, R. Canonical correlation analysis of video volume tensors for action categorization and detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 2009, 31 (8), 1415–1428.

64. Harandi, M. T.; Sanderson, C.; Wiliem, A.; Lovell, B. C. Kernel analysis over riemannian manifolds for visual recognition of actions, pedestrians and textures. WACV, 2012.

65. Lui, Y. M.; Beveridge, J. R.; Kirby, M. Action classification on product manifolds. In IEEE Conference on Computer Vision and Pattern Recognition, 2010.

66. Aggarwal, J. K.; Ryoo, M. S. human Activity Analysis: A Review; ACM Computing

Survey; The university of Texas, Austin, 2011.

67. Schuldt, C.; Laptev, I.; Caputo, B. Recognizing human actions: a local SVM approach. International conference on pattern Recognition, 2004; pp 32-36.

68. Cho, S. Y.; Byun, H. R. Human activity recognition using overlapping multi-feature descriptor. Science Direct ,Electronics Letters 2011, 47 (23).

69. Farid, A.; Modarres, A.; Soryani, M. Body posture graph: a new graph-based posture descriptor for human behaviour recognition. IET Comput. Vis. 2013, 7 (6), 488-499.

70. Sheng, B.; Yang, W.; Sun, C. Action recognition using direction-dependent feature pairs and non-negative low rank sparse model. Science Direct Neurocomputing 2015, 158, 73-80.

71. Arunnehru, J.; Geetha, M. K. Behavior recognition in surveillance video using temporal features. Fourth International Conference on Computing, Communications and Networking Technologies, 2013; pp 1-5.

72. Li, C.; Bailiang, S.; Liu, Y.; Wang, H.; Wang, J. Human action recognition using spatio-temoporal descriptor. 6th International Confrence on Image and Signal Processing, 2013; pp 107-111.

73. Zhang, Z.; Hu, Y.; Chan, S.; Chia, L. T. "Motion Context: A New Representation for Human Action Recognition. Proc. European Conf. Computer Vision, 2008; pp 817-829.