A THESIS REPORT

ON

HAND POSTURE RECOGNITION USING SKIN SALIENCY MAP AND TEXTURAL EVIDENCE

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIRNMENT FOR THE AWARD OF THE DEGREE OF

MASTER OF TECHNOLOGY

IN

SIGNAL PROCESSING AND DIGITAL DESIGN

SUBMITTED BY

PRIYADARSHANI

2K13/SPD/29

UNDER SUPERVISION OF

Dr. DINESH KUMAR VISHWAKARMA

ASSISTANT PROFESSOR

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY, DELHI (INDIA)

JUNE 2016

A THESIS REPORT

ON

HAND POSTURE RECOGNITION USING SKIN SALIENCY MAP AND TEXTURAL EVIDENCE

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIRNMENT FOR THE AWARD OF THE DEGREE OF

MASTER OF TECHNOLOGY

IN

SIGNAL PROCESSING AND DIGITAL DESIGN

SUBMITTED BY

PRIYADARSHANI

2K13/SPD/29

UNDER SUPERVISION OF

Dr. DINESH KUMAR VISHWAKARMA

ASSISTANT PROFESSOR

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY, DELHI (INDIA)

JUNE 2016

DECLARATION

I hereby declare that the work presented in this report, titled ""**Hand Posture Recognition Using Skin Saliency Map and Textural Evidence**", 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.

Priyadarshani

Roll No. 2K13/SPD/29

M.Tech Signal Processing and Digital Design

Date:

CERTIFICATE

This is to certify the research work embodied in this dissertation entitled "**Hand Posture Recognition Using Skin Saliency Map and Textural Evidence**" submitted by Miss Priyadarshani, Roll no. 2K13/SPD/29 student of Master of Technology in Signal Processing and Digital Design under Department of Electronics and 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.

(Dr. Dinesh K. Vishwakarma)
Assistant Professor
Dept. of Electronics and Communication Engineering
Delhi Technological University, Delhi, India.

Date:

ACKNOWLEDGEMENT

I express my sincere thanks and deep sense of gratitude to my guide, **Dr. Dinesh Kumar Vishwakarma**, Assistant Professor, Department of Electronics & Communication Engineering, Delhi Technological University, whose encouragement, the initial to 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, Department of Electronics & Communication Engineering, Delhi Technological University 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.

Priyadarshani

Roll No. 2K13/SPD/29

M.Tech Signal Processing and Digital Design

ABSTRACT

The aim of this thesis is to present a hand gesture recognition approach in static hand posture images. The major steps of the work includes (a) segmentation of hand region from rest of the image, (b) formation of saliency image (c) feature extraction using Gabor filter and Pyramid histogram of oriented gradients (PHOG). The YCbCr color model is used to detect the skin region of the hand, whereas the saliency map assigns a higher rank to the visually prominent area along with edges of the hand region. Gabor filter is used to extract texture feature at various orientations and scales while PHOG extract the shape of hand by computing the spatial distribution of skin saliency image. Finally, the extracted features are used to classify through Support Vector Machine (SVM). The performance of the proposed algorithm is demonstrated on publicly available datasets, and the recognition accuracy achieved on these datasets are compared with similar state-of-the-art, which shows superior performance.

Table of Contents

LIST OF FIGURES

LIST OF TABLE

CHAPTER 1

Hand Gesture Recognition System

1.1 INTRODUCTION

In this era of Digital world, day by day interaction between human and machine is becoming more vital, and how efficiently and precisely it works with least minimum time is one of the important concerns. Because of which importance of Hand gesture recognition is continuously growing for natural reason. Best way to interact with machine is by Visual interaction, which is without any physical contact, natural way of interface, easy and effective. So in Visual pattern analysis, Hand gesture recognition is area of research which is having application in different areas like sign language recognition, human-robot interaction (HRI), human-computer interaction (HCI), Smart interactive television and virtual reality (VR) [1]. This potential of Human vision system galvanized the progression of computational models of human vision system. But there is difficulty in recognition of hand gestures due to presence of cluttered and complex background. Human visual system is difficult to understand since it rapidly and conveniently identifies numerous other number of objects in cluttered backgrounds, natural scenes and particular patterns. There are four different phases to identify the gesture in Hand gesture recognition system. First being data acquisition, second Object segmentation and pre-processing, third feature extraction and fourth the recognition. For a hand gesture recognition system where our object is hand, first the hand image is taken by suitable input device

then the hand or object is segmented from the image to identify the hand from the cluttered background and other parts of body, after it; pre-processing is done on the image to remove noises, to identify edges/ contours, normalized to develop the desired model. For recognition the features or properties of the hand gesture are calculated from the segmented or pre-processed image. The flow diagram of Hand gesture recognition system is shown in figure 1:

Figure 1.1 Flow diagram of Hand gesture recognition system

1.2 IMAGE INPUT

An image is defined in a two dimensional space by a function f(x, y), where x and y are plane or spatial coordinates, and at any pair of coordinates (x, y) the amplitude represent the intensity or grey level at that point of the image [2]. When the value of intensity level, x and y all are finite, discrete then the image is called digital image.

The input device for an input image can be either a sensor such as Kinect sensor for depth images or a camera depending on application requirement. There are four basic types of digital images: binary, grayscale, true color or RGB and indexed.

In binary image each pixel value is black or white (0 or 255). In greyscale image pixel value ranges from 0 to 255 or we can say with a shades of grey. In RGB image each pixel has certain color, decided by amount of red, green and blue, which lies in the range of 0 to 255.

1.3 IMAGE SEGMENTATION

The task of object segmentation remain as one of the primary, important and challenging tasks in image processing applications such as Medical, Object detection, Hand gesture recognition, Facial expression recognition, Object tracking, Traffic monitoring, video surveillance etc. to extract information from a certain image. The extraction of the object needs to be robust and reliable so that the demands of application could be meet. In each of the Image processing application to detect, track or recognize the object in an image we first need to separate the particular object of interest or we can say foreground from the rest of the unnecessary objects or we can say background. The separation of foreground from background is known as Object Segmentation. There are various techniques for the task of object segmentation. Which technique needs to be used depends on the problem or we can say the task of object segmentation is problem specific. Images are basically divided into two types: gray scale and color image. The object segmentation technique used for a gray scale image is totally different from that of a color image. The segmentation is based on measurements taken from the image and might be grey level, color, texture, depth or motion [3]. The basic techniques still used by researchers for object segmentation are Edge Detection, Threshold, Histogram, Region based methods, and Watershed Transformation [4]. Some of the most famous object segmentation techniques including Edge based segmentation, Fuzzy theory based segmentation, Partial Differential Equation (PDE) based segmentation, Artificial Neural Network (ANN) based segmentation, threshold based object segmentation, and Region based image segmentation [3] are highlighted in figure 2.

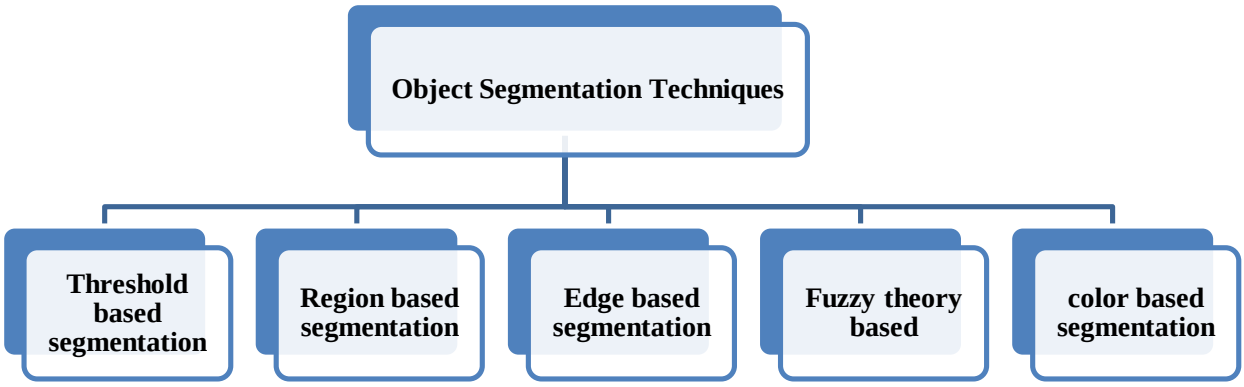


Figure 1.2: Various Image Segmentation Techniques

The most popular techniques used for the segmentation are explained in the subsequent sections.

1.3.1 Threshold based segmentation

Threshold based segmentation technique is a longstanding, simple and famous method for image segmentation. Object segmentation by setting a threshold level is a simple but powerful method for segmenting images which have light objects but dark background. By thresholding operation a multilevel image is converted into a binary image. It chooses a proper threshold level 'T' and divides image pixels into numerous regions and isolate objects from the background. Any pixel (x, y) in the image is assumed as a part of the object if its intensity level is higher or equal to threshold level 'T' i.e., f(x, y) ≥T, otherwise it belongs to the background. Basically, two types of thresholding approaches are defined: global and local thresholding. If the threshold level 'T' is constant then it is identified as global thresholding, else it is local thresholding. When the background illumination is irregular global thresholding approach may fail. To compensate for irregular illumination in local thresholding several thresholds are introduced. Thresholding method has certain disadvantages. It only creates two classes, thus

it is not useful for multichannel images. Thresholding is sensitive to noise as it does not take spatial features of the image into account which in turn degrades histogram of an image and makes separation more difficult.

1.3.2 Region-based segmentation

Region-based segmentation techniques are based on the point that a pixel could not be assumed as a part of the object based merely on its gray scale value (as threshold-based methods do). Degree of connectivity among pixels is incorporate in this method in order to decide either these pixels fit to the same object (or region) or not. Mathematically, region-based segmentation technique can be defined as a systematic manner to divide the image *I* into *n* objects (or regions), $R_1, R_2, \ldots\ldots R_n$ such that, $\sum_{i=1}^{n} R_i = I$, Where R_i is a connected region and i =1, 2, …., n. Each pixel will be categorized as fitting to one of the *n* regions in the image. The regions cannot overlap and a certain criterion needs to be fulfilled so that a pixel is allowed to belong to a certain region, i.e. all pixel values in a region must be within a range of intensities. Finally, two adjacent regions cannot be similar. In Region based segmentation, we have to find homogeneous regions according to a specific criterion (intensity value, texture) [3]. There are two methodologies in region-based methods:

- Region Growing
- Region Splitting and Merging.

In Region Growing Method, clusters of neighboring pixels of a region are made that verify specific assumptions. A Seed region is initialized and expanded to include all identical neighbors and the procedure is repeated. The development of region growing ends when no pixel is left to be categorized. In region splitting method, the image is Split into *n* number of regions based on a specific

assumption. The procedure starts with the whole image as a seed. If the seed is non-uniform then it is divided into fixed number of sub-regions, usually four. The region splitting procedure is repeated considering each sub-region in the image as a seed. The procedure finishes when all sub-regions become uniform. In Region Merging Method, any neighboring regions that are homogenous enough are merged.

1.3.3 Edge Based Image Segmentation

In object segmentation edge detection of the object is one of the key application. Edge-based segmentations rely on edges found in an image by edge detecting operators—these edges mark image locations of discontinuities in gray level, color, texture, etc. Image resulting from edge detection cannot be used as a segmentation result. Supplementary processing steps must follow to combine edges into edge chains that correspond better with borders in the image. The most common problems of edge-based segmentation are an edge presence in locations where there is no border, and no edge presence where a real border exists (false alarms and missed detections). The process of partitioning a digital image into multiple regions or sets of pixels is called image segmentation. Edge is a boundary between two homogeneous regions. Edge detection refers to the process of identifying and locating sharp discontinuities in an image. Edge detection is one of the most frequently used techniques in digital image processing. The boundaries of object surfaces in a scene often lead to oriented localized changes in intensity of an image, called edges. Edge detection techniques transform images to edge images benefiting from the changes of grey tones in the images. Edges are the sign of lack of continuity, and ending. As a result of this transformation, edge image is obtained without encountering any changes in physical qualities of the main image. An Edge in an image is a

significant local change in the image intensity, usually associated with a discontinuity in either the image intensity or the first derivative of the image intensity. Discontinuities in the image intensity can be either Step edge, where the image intensity abruptly changes from one value on one side of the discontinuity to a different value on the opposite side, or Line Edges, where the image intensity abruptly changes value but then returns to the starting value within some short distance.

1.3.4 Color based segmentation

In hand gesture system skin color segmentation is an important task as the color of hand is of skin color. The purpose of skin color model is to discover either a pixel falls in the range of skin color or in non-skin color. Previous studies have suggested that the skin color of human is not reliant on human race or on the wavelength of the visible light. This observation visibly informs the requirement of elimination of the luminance feature in appropriate way by the color space model. Pure color information need to be attained and it should be not reliant on the illumination of the scene. All the color space models are mathematical depiction of a set of colors. All the color space models are extracted from the RGB color space model information is provided by input devices such as cameras. The most widespread color space models are YCbCr, HSV, HSI. The HSV (hue, saturation, value) color space is developed to be more intuitive in manipulating color and designed to approximate the way humans perceive and interpret color. The HSV color space is preferred for manipulation of hue and saturation i.e. to shift color or adjust the amount of color since it yields a greater dynamic range of saturation.

The HSI (hue, saturation and intensity) is similar to HSV model. The main difference between these two models is the computing of the brightness component (I and V), which determines the distribution and dynamic range of both the brightness and saturation. The HSI method is best color space for the traditional image processing function like Convolution, Equalization, and Histogram. The YCbCr is another color space unlike the RGB color space, here the luminance or brightness or intensity is separated from the chrominance or pure color value. The value of Y represents the luminance value and Cb and Cr represents the color or chrominance value, these are also known as color difference of the image.

There are studies in literature [5] [6] to extract significant skin color boundaries to segment skin pixels in a given image effectively. These boundaries were designed to include all possible skin color values and named as skin color model. For HSV color space, a pixel is classified as skin color if the conditions given in Eq. 1.1 and 1.2 are satisfied otherwise non skin color.

$$0 < H < 52 \tag{1.1}$$

$$0.21 < S < 0.70 \tag{1.2}$$

For YCbCr color space, a pixel is classified as skin color if the following conditions are satisfied otherwise non skin color.

$$79 < Y \tag{1.3}$$

$$83 < Cb < 135 \tag{1.4}$$

$$132 < Cr < 180 \tag{1.5}$$

For the segmentation of skin region from hand posture can be done using the values given in Eq. 1.4 and 1.5. The values of Y is not considered due to its instability and sensitive to illumination change.

1.4 FEATURES

For any classification problem extracted features should be unique so that the objects can be easily distinguished in the feature space. Mainly in Hand gesture recognition three features are considered: Shape, color and texture. Based on application or requirement any of the features or combination of two or all the three of them can be calculated and used for recognition.
Different visual features are as explained below:

Color: The color of a particular object is influenced largely by two physical elements, illuminant's spectral power distribution and object's surface reflectance properties. In image processing, usually RGB color space is used to represent color. However, the RGB space is not a uniform color space, that is, the differences between the colors in the RGB space do not correspond to the color differences perceived by humans [7]. Furthermore, the RGB magnitudes are highly correlated. Compared to RGB space, L$*u*v*$ and $L*a*b*$ are uniform color spaces, while HSV (Hue, Saturation, Value) is an approximately uniform color space. Though, these color spaces are quite sensitive to noise. At last, no color space is more efficient than other it all depends on application and requirement, hence a multiple color spaces are used in recognition.

Edges: In any image edges represents the abrupt change in intensity level. Edge detection is used to identify these changes. The most attractive feature of edge based features are that edges are very less sensitive to noise. Commonly used edge detection technique is the Canny edge detector because of its simplicity and accuracy.

Texture: Texture gives extent of intensity variation in a surface which measures properties such as smoothness and regularity. Compared to color, texture requires

a processing step to generate the descriptors. There are various texture descriptors. Similar to edge features, the texture features are less sensitive to illumination changes compared to color.

Feature selection is closely related to the object representation. For example, color is used as a feature for histogram-based appearance representations, while for contour-based representation, object edges are usually used as features.

1.5 FEATURE SELECTION

After the features have generated the most significant features are selected for further processing. All the generated features can't be used for further processing because,

- More number of features makes system more complex
- It takes greater time while execution.

The most common methods used for feature reduction or to reduce the dimension of input data is **PCA** (Principal Component Analysis) and **LDA** (Linear Discriminant Analysis).

1.5.1 Principal Component Analysis (PCA)

This technique projects the data onto the direction that have largest variance. PCA [8] reduces the dimension of input data by identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. The main advantage of PCA is that it reduces the dimension of input data without much loss of information. PCA can be performed in the following steps:

- Get input data vector (X).

- Find the mean subtracted data (X - μ).
- Calculate the covariance matrix (∑).
- Calculate the eigenvectors and eigenvalues of the covariance matrix.
- Choose the eigenvectors corresponding to the most significant eigenvalues and form the feature vector by constructing a matrix of selected eigenvalues.
- Feature Vector = (eigenvector1, eigenvector…… eigenvectorN)
- Derive the new data set as follow :

Final Data = Row Feature Vector x Row Data Adjust

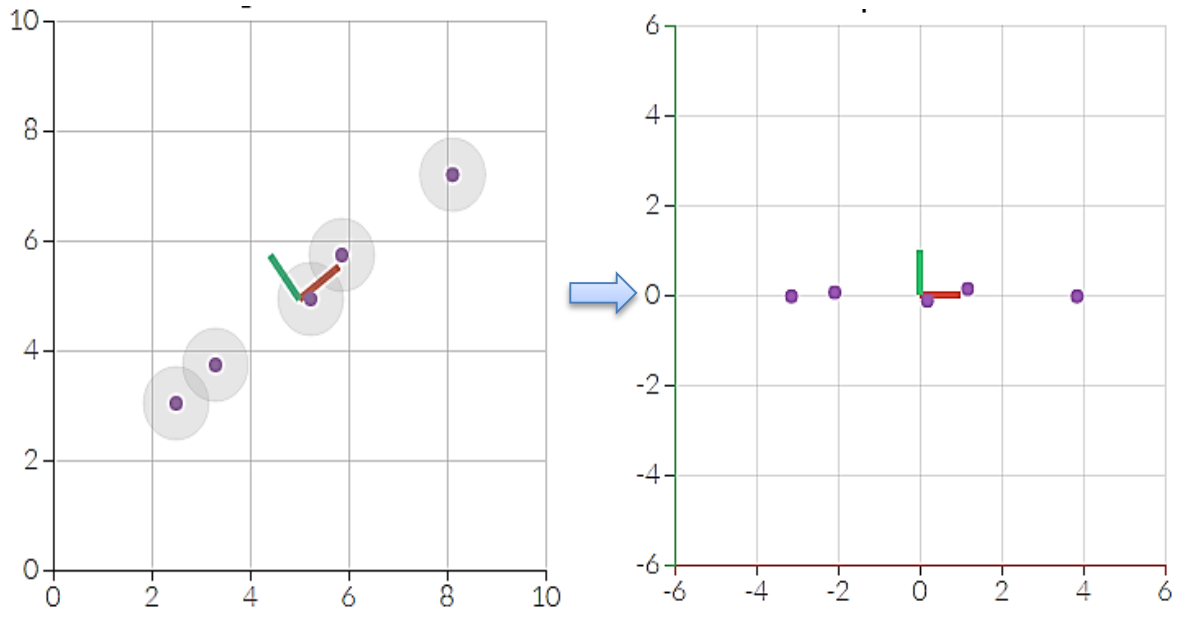
Where,

Row Feature Vector = $(\text{Feature Vector})^T$

Row Data Adjust = $(X - \mu)^T$

- Original data set can be derived back as :

Row Original Data = $(\text{Row Feature Vector}^T \times \text{Final Data}) + \text{Original Mean}$



(a) Data set dependent on two dimensions (b) Data set dependent on one dimension

Figure 1.3: Principal Component Analysis [8]

1.5.2 Linear Discriminant Analysis (LDA)

LDA projects a d-dimensional input data to a smaller subspace of k dimensions (k < d) while preserving the class-discriminatory information. A good projection vector is that which maximizes the separation between the projections thus the objective function can be defined as given in Eq. 1.6 (for two class problem):

$$J(w) = |\tilde{\mu}_1 - \tilde{\mu}_2| = |w^T(\mu_1 - \mu_2)| \quad (1.6)$$

Fisher modifies the objective function by considering the within class scatter. Thus for Fisher's LDA objective function [10] is as given in Eq. 1.7:

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \quad (1.7)$$

Where,

S_B = between class scatter matrix = $(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$

S_W = within class scatter matrix = $S_1 + S_2$

After maximization, the optimal value of 'w' can be given as:

$$W^* = \text{argmax}\left\{\frac{w^T S_B w}{w^T S_W w}\right\} = S_w^{-1}(\mu_1 - \mu_2) \quad (1.8)$$

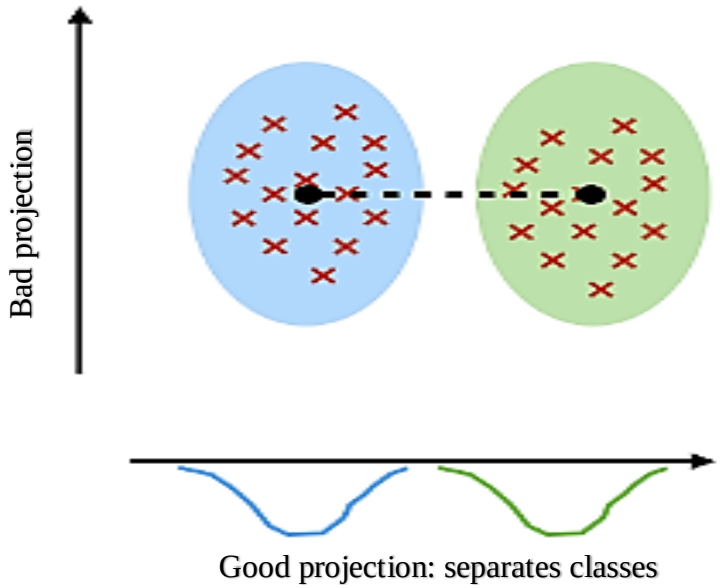


Figure 1.4: Linear Discriminant Analysis [8]

1.6 Feature Classification

Once we get the features from feature reduction technique then comes the last but the most important task of feature classification, using which classification among various classes is achieved. There are two methods for pattern classification: supervised and unsupervised classification.

1.6.1 Supervised classification:

The supervised classification of input data in the pattern recognition method uses supervised learning algorithms that create classifiers based on training data from different object classes. The classifier then accepts input data and assigns the appropriate object or class label. Supervised learning includes two categories of algorithms: Classification and Regression**.** Classification is used for categorical response values, where the data can be separated into specific "classes" while regression is used for continuous-response values.

Common classification algorithms include:

- Support vector machines (SVM)
- Neural networks
- Naïve Bayes classifier
- Decision trees
- Discriminant analysis
- Nearest neighbors (*k*-NN)

Common regression algorithms include:

- Linear regression
- Nonlinear regression

- Generalized linear models
- Decision trees
- Neural networks

1.6.2 Unsupervised classification:

The unsupervised classification method works by finding hidden structures in unlabeled data using segmentation or clustering techniques. The most common unsupervised learning method is cluster analysis, which is used for exploratory data analysis to find hidden patterns or grouping in data. The clusters are modeled using a measure of similarity which is defined upon metrics such as Euclidean or probabilistic distance.

Common unsupervised classification methods include:

- K-means clustering
- Gaussian mixture models
- Hidden Markov models

Support Vector Machine (SVM)

A Support Vector Machine (SVM) [9] is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (*supervised learning*), the algorithm outputs an optimal hyperplane which categorizes new examples. In this algorithm, each data item is plotted as a point in n-dimensional space (n = number of features) with the value of each feature being the value of a particular coordinate. Then, classification is done by finding the hyper-plane that differentiate the two classes very well.

Hyperplane can be represented by equation 1.9:

$$wx + b = 0 \tag{1.9}$$

Where, w = weight vector normal to hyperplane and b = bias or threshold

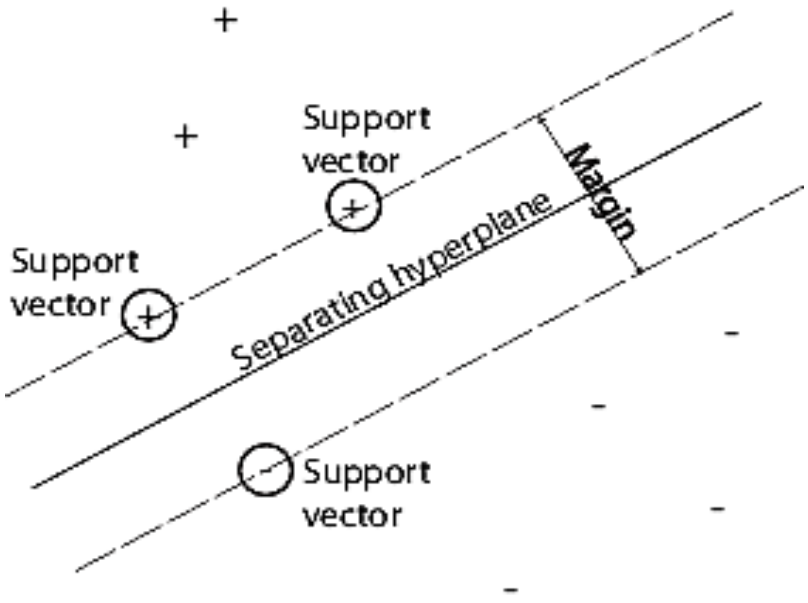


Figure 1.5: Support Vector Machine

- **Linear-SVM classifier**

Let the simplest case, in which the training patterns are linearly separable. That is, there exists a linear function of the form by Eq. 1.10:

$$f(x) = w^T x + b \tag{1.10}$$

Such that for each training example x_i, the function yields

$$f(x_i) \geq 0 \; for \; y_i = +1$$

$$f(x_i) < 0 \; for \; y_i = -1$$

Optimal hyperplane can be found by minimizing the cost function in Eq. 1.11

$$J(w) = \frac{1}{2} w^T w \tag{1.11}$$

Subject to the separability conditions given in Eq. 1.12

$$wx_i + b \geq +1 \; for \; y_i = +1$$

$$wx_i + b \leq -1 \; for \; y_i = -1 \; for \; i = 1, 2, 3 \ldots \ldots L \tag{1.12}$$

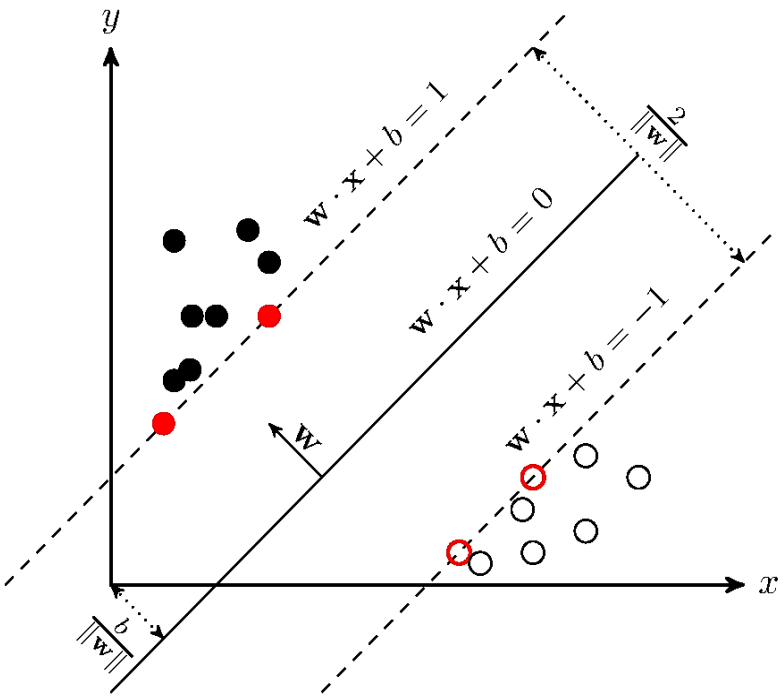


Figure 1.6: Linear-SVM

- **Non-Linear SVM classifier**

  The linear SVM can be extended to a non-linear classifier using a non-linear operator $\emptyset(.)$ to map the input pattern X into higher dimensional space H. The non-linear SVM classifier so obtained is given in Eq. 1.13:

$$f(x) = w^T \emptyset(x) + b \tag{1.13}$$

Which is linear in terms of the transformed data φ(x) but non-linear in terms of the original data x ϵ R^n

Now optimal plane can be found by minimizing the function in Eq. 1.14

$$\min J(w, \xi) = \frac{1}{2} w^T w + c \sum_{i=1}^{l} \xi_i \tag{1.14}$$

Subject to the constrain given in Eq. 1.15,

$$w^T \emptyset(x_i) + b \geq 1 - \xi_i \tag{1.15}$$

$$\xi_i \geq 0 \; ; i = 1,2,3 \ldots \ldots . L \geq 0$$

Kernel functions are used to map the low dimensional data into the high dimensional feature space where data points are linearly separable. Different kernel functions for SVM are:

I. Polynomial function of degree 'd' $K(x, y) = (\tau + x^T y)d$

II. Radial basis function $K(x, y) = \exp(-||x - y||^2/\sigma^2)$

III. tan-sigmoid function $K(x, y) = tanh(k_1 x^T y + k_2)$

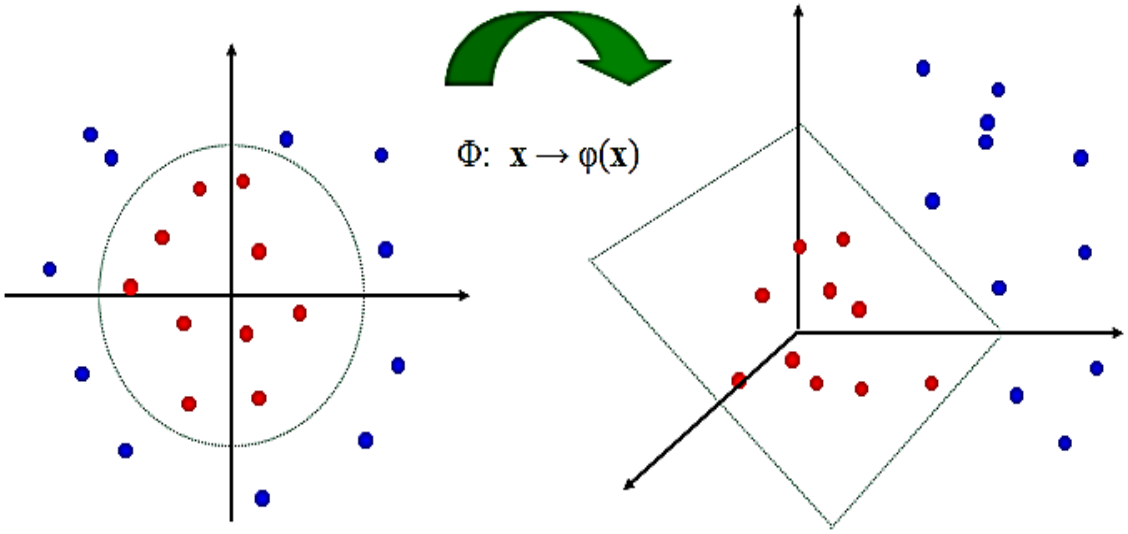


Figure 1.7: Transformation of non-linear SVM into linear-SVM

- **Multi-Class SVM**

  There are two approaches to multi-class SVM:

  - One against the rest approach
  - One against the one approach

  1. One against the rest approach

  One against all constructs k SVM models where k is the number of classes. The m^{th} class with positive labels, and all other examples with negative labels. Thus given 'l' training data $(x_1, y_1) \ldots\ldots (x_l, y_l)$ where $x_i \in R^n$, $i =$

$1 \ldots\ldots l \; and \; y_i \in \{1, 2 \ldots\ldots k\}$ is the class of x_i, the m^{th} SVM solves the following problem:

$$\min(w^m, b^m, \xi^m) \frac{1}{2}(w^m)^T w^m + c \sum_{i=1}^{l} \xi_i^m \tag{1.16}$$

$$(w^m)^T \emptyset(x_i) + b^m \geq 1 - \xi_i^m \; if \; y_i = m_i$$

$$(w^m)^T \emptyset(x_i) + b^m \leq -1 + \xi_i^m \; if \; y_i \neq m_i$$

$$\xi_i^m \geq 0 \; for \; i = 1, 2 \ldots\ldots l \geq 0$$

Where the training data x_i are mapped to a higher dimensional space by the function Φ and c is the penalty parameter.

Solving the above equation x is defined to the class which has largest value of the decision function.

$$Class \; of \; x = argmax_{m=1,2\ldots..k}((w^m)^T \emptyset(x) + b^m \tag{1.17}$$

Drawbacks of One against the rest approach

- Memory requirement is very high
- Training sample size is unbalanced

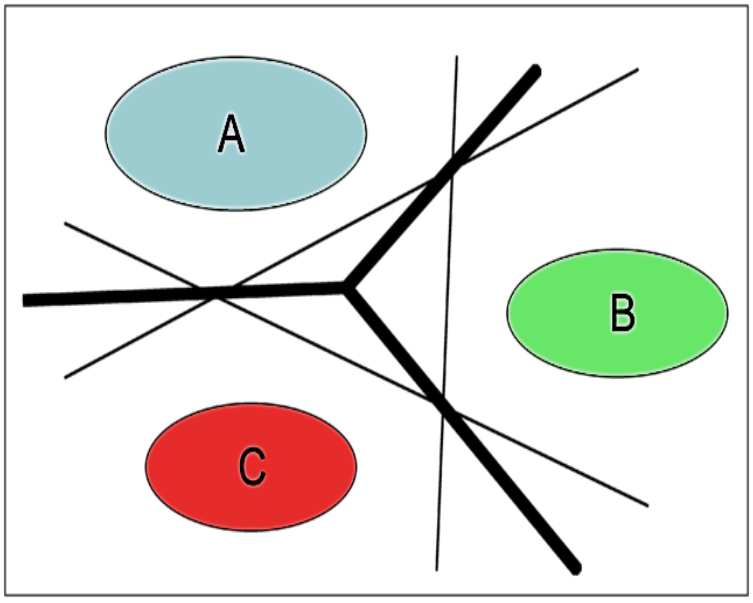


Figure 1.8: One against the rest approach of SVM [10]

2. One against one approach

One against one method constructs $k(k-1)/2$ classifiers where each one is trained on data from two classes. For training data from the i^{th} and j^{th} class, we solve the following binary classification problem

$$\min\left(w^{ij}, b^{ij}, \xi^{ij}\right) \frac{1}{2}(w^{ij})^T w^{ij} + c\sum_t \xi_t^{ij} \tag{1.18}$$

$$(w^{ij})^T \emptyset(x_t) + b^{ij} \geq 1 - \xi_i^{ij} \; if \; y_t = i$$

$$(w^{ij})^T \emptyset(x_t) + b^{ij} \leq -1 + \xi_i^{ij} \; if \; y_t = j$$

$$\xi_t^{ij} \geq 0 \; for \; i = 1,2 \ldots \ldots l$$

If $sign((w^{ij})^T \emptyset(x_t) + b^{ij})$ says x is in i^{th} class, vote of the i^{th} class is added by one otherwise vote of j^{th} class is increased by one. Then we predict x is in class with largest vote.

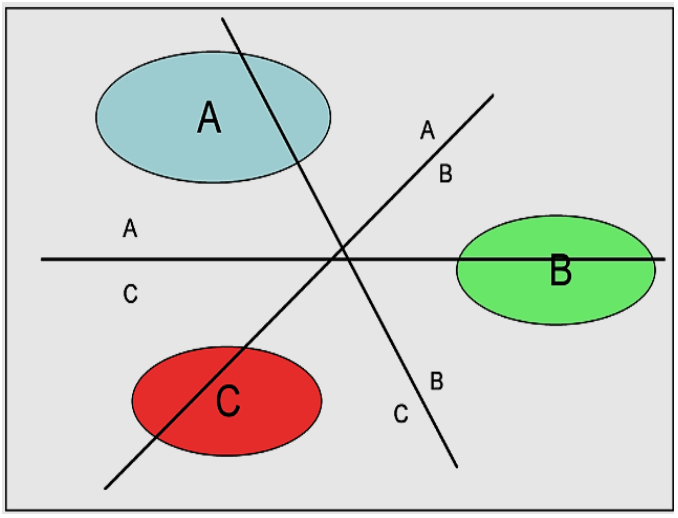


Figure 1.9: One against one approach of SVM [10]

Drawback of one against one approach: Increase in the number of classifiers as number of class increases.

1.7 THESIS OUTLINE:

The outline of this thesis is as follows.

Chapter 1: It explains the steps involved in

Chapter 2: This chapter focuses on the Literature Review of Object Segmentation and Texture Information.

Chapter 3: This chapter explains the steps of proposed methodology.

Chapter 4: In this chapter, the experimental results obtained for three dataset are compared and discussed.

Chapter 5: In this chapter, conclusion of the thesis is given and the future work is discussed.

CHAPTER 2

LITREATURE REVIEW

Hand gesture recognition is an area of research from quite long time. Scholars have progressed to a level that hand gesture is now used into practical usage in real time devices. For example, Samsung A7 uses hand gesture to click a selfie, a very basic application in daily usage. Though controlling a device wholly using hand gesture have not been possible yet or we can say the potential of this technology is yet to be tapped.

Under this chapter we will review the work done until now by various scholars. The various methods used for hand gesture recognition by scholars are:

- Hand Gesture Recognition Using Kinect Sensor.
- Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques.
- Hand gestures recognition using dynamic Bayesian networks.
- Gloved and Free Hand Tracking based Hand Gesture Recognition.
- HCI for Smart Environment Applications Using Fuzzy Hand Posture and Gesture Models
- A Static Hand Gesture Recognition Algorithm Using K-Mean Based Radial Basis Function Neural Network.
- Hidden Markov Model - based Gesture Recognition with Overlapping Hand-Head/Hand-Hand Estimated using Kalman Filter.

- Hand Gesture Recognition based on Shape Parameters.

2.1 Hand Gesture Recognition Using Kinect Sensor

Ren et al. [11] has proposed, a novel distance metric, Finger-Earth Mover's Distance, to measure the dissimilarity between hand shapes to handle the noisy hand shapes obtained from the Kinect sensor. It matches only finger parts not the whole hand and it can better distinguish hand gestures of slight differences. It gives 93.2% mean accuracy on a challenging 10-gesture dataset and is very efficient, robust to hand articulations, distortions and orientation or scale changes, and can work in uncontrolled environments (cluttered backgrounds and lighting conditions).

Due to low resolution of Kinect depth map, of only 640 × 480, it is difficult to detect and segment a small object from an image with this resolution, e.g., human hand occupies a very small portion of the image with more complex articulations. Thus segmentation of the hand is usually inaccurate, thus may significantly affect the recognition step. They have implemented the proposed technique in two real-life HCI applications on top of our hand gesture recognition system: Arithmetic computation and Rock-paper-scissors game.

2.2 Bag-of-Features and Support Vector Machine Techniques

Dardas & Georganas [12] propose a real-time hand gesture detection and recognition using Bag-of-Features and Support Vector Machine Techniques. In this a bare hand detected and tracked in cluttered background using skin detection and hand posture contour comparison algorithm after face subtraction, the hand gestures are recognized via bag-of-features and multiclass support vector machine (SVM) and building a grammar that generates gesture commands to

control an application. In training stage, after extracting the key points for every training image using the scale invariance feature transform (SIFT), a vector quantization technique will map key points from every training image into a unified dimensional histogram vector (bag-of-words) after K-means clustering, histogram is treated as an input vector for a multiclass SVM to build the training classifier.

It gives satisfactory real-time performance regardless of the frame resolution size as well as high classification accuracy of 96.23% under variable scale, orientation and illumination conditions, and cluttered background. The accuracy of the result is affected by the quality of the webcam in the training and testing stages, the number of the training images, and choosing number of clusters to build the cluster model.

2.3 Hand gestures recognition using dynamic Bayesian networks

Shiravandi et al. [13] study includes two main subdivisions namely: hand posture recognition and dynamic hand gesture recognition (without hand posture recognition). In the first session, after hand segmentation using a method based on histogram of direction and fuzzy SVM classifier, we train the posture recognition system. In the second session, after skin detection and face and hands segmentation, their tracing were carried out by means of Kalman filter. Then, by tracing the obtained data, the positions of hand is achieved. For combining the achieved data and output of hand posture recognition unit they utilize Bayesian dynamic network. For recognition of 12 hand gestures in this study, 12 Bayesian dynamic networks with two distinct designs are used. The difference between these two models is in the utilizing features and their relations with each other.

Therefore, one of these models is used based on each gesture feature. The results of implementation show the about 90% average accuracy for all gestures.

2.4 HCI for Smart Environment Applications using Fuzzy Hand Posture and Gesture Models

Várkonyi-Kóczy & Tusor [14] introduce, a hand posture and gesture modeling and recognition system, which can be used as an interface to make possible communication with smart environment (intelligent space) by simple hand gestures. The system transforms preprocessed data of the detected hand into a fuzzy hand-posture feature model by using fuzzy neural networks and based on this model determines the actual hand posture applying fuzzy inference. Finally, from the sequence of detected hand postures, the system can recognize the hand gesture of the user.

This system works with six predefined hand postures and gestures consisting any series composed of these six postures. The correct identification rate proved to be in average above 96%. The number of used hand postures can be considered little, compared with the usual number of elements of the traditional sign languages. Limitation is that we can use only a homogenous background. This is possibly the most limiting factor of the application.

2.5 HMM based Gesture Recognition with Overlapping Hand-Head/Hand-Hand Estimated using Kalman Filter

Gaus & Wong [15] propose a HMM based Gesture Recognition. First, we apply skin segmentation procedure throughout the input frames in order to detect only skin region. Then, we proceed to feature extraction process consisting of centroids, hand distance and hand orientation collecting. Kalman Filter is used to

identify the overlapping hand-head or hand-hand region. After having extracted the feature vector, the hand gesture trajectory is represented by gesture path in order to reduce system complexity. We apply Hidden Markov Model (HMM) to recognize the input gesture. The gesture to be recognized is separately scored against different states of HMMs. The model with the highest score indicates the corresponding gesture. The recognition rate is about 83%.

2.6 Hand Gesture Recognition based on Shape Parameters

Panwar [16] propose a hand gesture recognition based on shape parameters**.** In this the input sequence of images through web cam it uses some pre-processing steps for removal of background noise and employs K-means clustering for segmenting the hand object from rest of the background, so that only segmented significant cluster or hand object is to be processed in order to calculate shape based features. Strength: simplicity, ease of implementation, and does not required any significant amount of training or post processing, provide with the higher recognition rate with minimum computation time.

Weakness: Need to define certain parameters and threshold values experimentally since it does not follow any systematic approach for gesture recognition, and maximum parameters taken in this approach are based on assumption made after testing number of images.

2.7 Static Hand Gesture Recognition Algorithm Using K-Mean Based Radial Basis Function Neural Network

Ghosh & Ari [17] has given a novel technique in this paper to obtain a rotation invariant gesture image, by coinciding the 1st principal component of the segmented hand gestures with vertical axes. A localized contour sequence (LCS)

based feature is used here to classify the hand gestures. A k-mean based radial basis function neural network (RBFNN) is also proposed here for classification of hand gestures from LCS based feature set. The proposed k-mean based RBF neural network yields 99.6% accuracy for classification of 500 gesture image database and shows better performance compared to MLP-BP Neural Network as reported in earlier research work.

From these earlier state-of-the-art methods, it has been observed that the following issues in the above hand gesture recognition systems limit the recognition rate, that are removed in our state-of-the-art method. The limitations possessed by the various methods are explained below:

- Due to low resolution of Kinect depth map, of only 640 × 480, it is difficult to detect and segment a small object from an image with this resolution, e.g., human hand occupies a very small portion of the image with more complex articulations. Thus segmentation of the hand is usually inaccurate, thus may significantly affect the recognition accuracy.
- In Smart Environment Applications Using Fuzzy Hand Posture and Gesture Models, the number of used hand postures can be considered less, compared with the usual number of elements of the traditional sign languages. Limitation is that we can use only a homogenous background. This is possibly the most limiting factor of the application.
- The detection accuracy achieved through HMM based Gesture Recognition system is quite low and time taken to detect a single posture is high due to complexity of HMM.
- HGR system based on Shape Parameters do not follow any systematic approach. Maximum parameters taken based on assumption.

CHAPTER 3

PROPOSED METHODOLOGY

In this work, first segmentation of hand posture is done using a combined method of skin segmentation and saliency map to remove the complex background and then texture feature is extracted using Gabor filter [18] and shape feature using Pyramidal Histogram of oriented Gradient (PHOG) [19]. Finally the classification among various classes is carried out using multi class Support Vector Machine (SVM). The workflow diagram of the proposed model is given below in Figure 3.1:

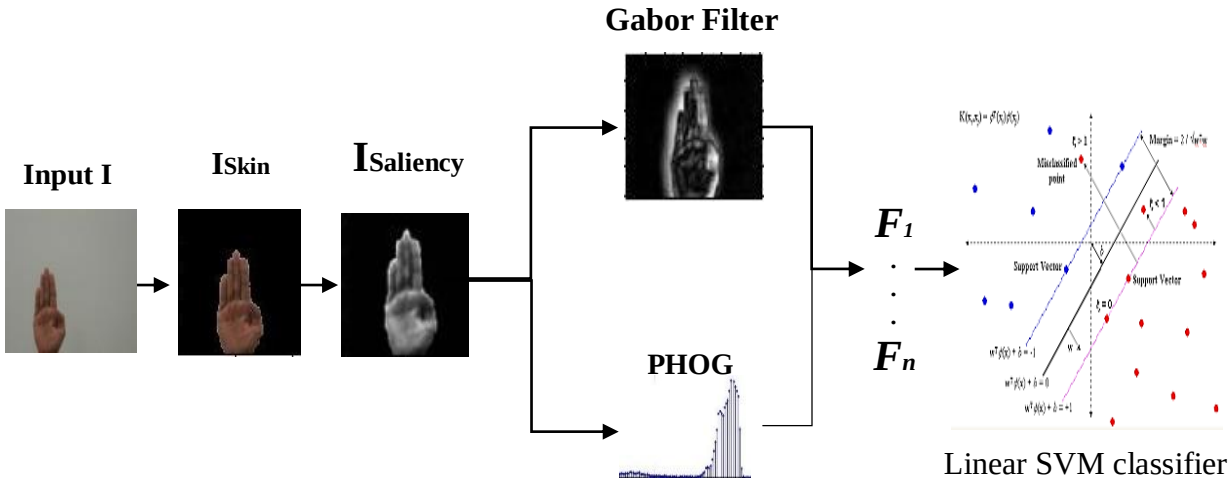


Figure 3. 1: The framework of the proposed hand posture detection and recognition model.

3.1 Hand Posture Segmentation

For the extraction of hand posture from rest of the image, a color skin based segmentation approach is used. In this approach, two important steps are

performed. First the skin region is marked in hand image using YCbCr color model and then saliency map is obtained of the resulting hand gesture images. The details of the two methods are described in the following sections.

3.1.1 Color Segmentation:

The aim of this step is to remove the non-skin color component from the input image or to detect the skin color region in the input image. In YCbCr the luminance or brightness or intensity is separated from the chrominance or pure color value, Y represents the luminance value, Cb and Cr represents the color or chrominance value, these are also known as color difference of the image. By using the histogram method it has been found that a skin-color region can be identified by the presence of a certain set of chrominance (i.e, Cr and Cb) values narrowly and consistently distributed in the YCbCr color space. For YCbCr color space, a pixel is classified as skin color if $77<Cb< 127$ and $133< Cr< 173$ otherwise it would be considered non skin color [20] [21].

3.1.2 Saliency Map Generation:

The output of YCbCr segmented image is not very good as it contains noise along with it, thus to suppress the effect of noise and segment the hand from the rest of the background we have taken saliency maps of the YCbCr segmented image. The Saliency maps [22] used in this method assigns higher rank not just to the edges but also to the visually prominent entire area. While in other methods used previously [23] [24] gives higher energy merely at edges of the objects. This method uniformly assigns saliency value to entire salient region and thus preserve the boundaries of the region. The drawbacks of saliency map used previously [23] [24] are removed by calculating global saliency of pixel rather than local

saliency of pixel or in other words, by uniformly assigning saliency values to whole salient area, rather than just boundaries or texture areas. In previous methods it was measured in terms of only intensity while here both intensity and color terms are measured. The saliency map [22] is attained by estimating the Euclidean distance between the average *LAB* vector value of an input image and each pixel of a Gaussian blurred version of the same input image:

$$E_{LAB}(X,Y) = \left\| J_{\mu} - J_{m\times m}(X,Y) \right\| \tag{3.1}$$

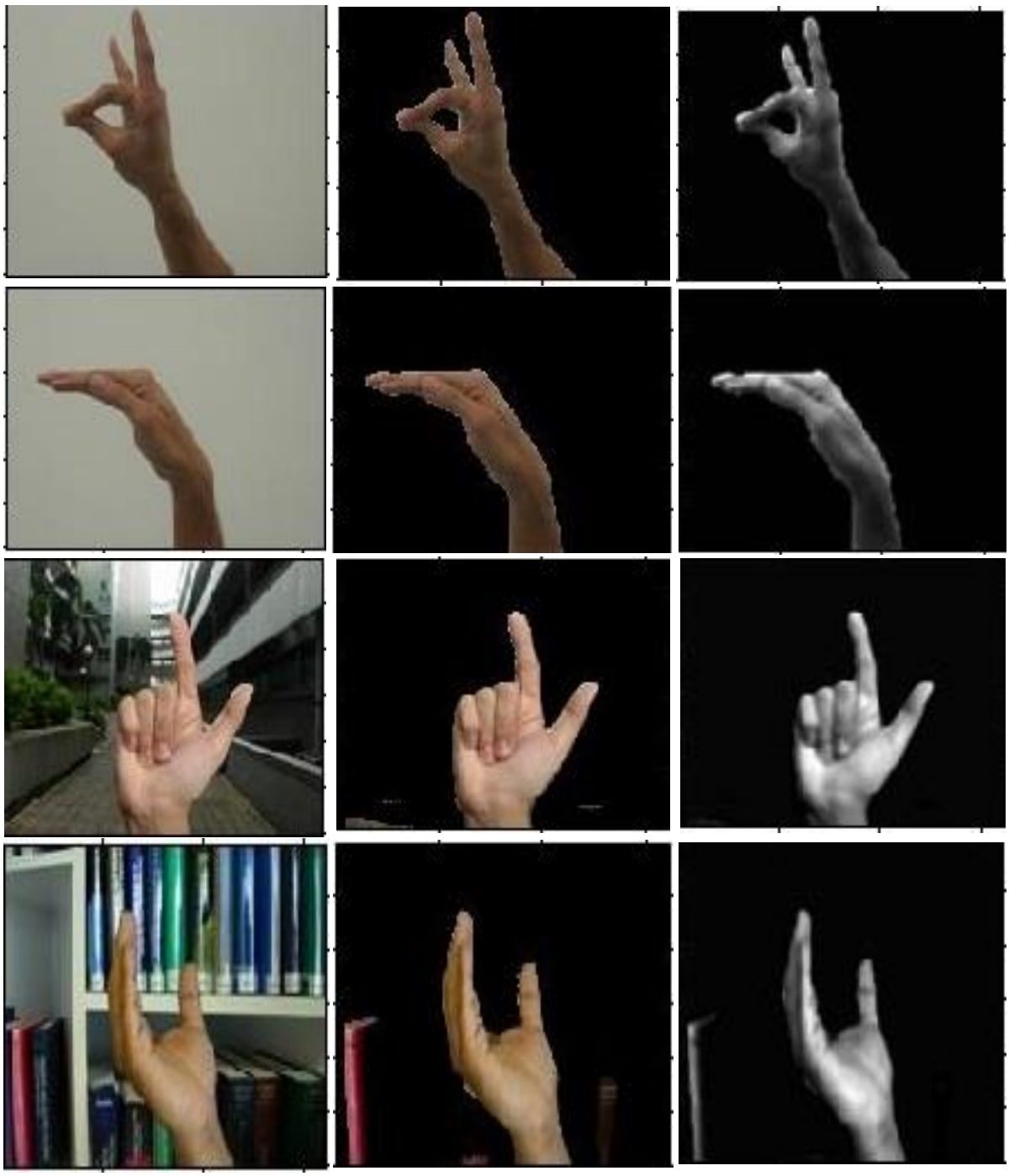
Where $E_{LAB}(X,Y)$ is the value of pixel saliency at location (X, Y), $\mathbf{J}_{\mu}$ is the average of entire *LAB* pixel vectors of the image, $\mathbf{J}_{m\times m}(X, Y)$ is the corresponding image pixel vector value in the Gaussian blurred version of the original image, and $\|\quad\|$ represent the Euclidean distance in *LAB* color space also known as L_2 norm. The *LAB* color space is used because Euclidean distances in this particular color space are almost perceptually uniform. Saliency maps are generated using Eq. 3.1 coupled with the modified forward energy terms Eq. 3.2 to overcome the limitations of saliency maps used previously by re-targeting schemes [23] [24].

$$C_u(X,Y) = \|J(X+1,Y) - J(X-1,Y)\| \tag{3.2}$$

$$C_l(X,Y) = \|J(X,Y-1) - J(X-1,Y)\| + C_u(X,Y) \tag{3.3}$$

$$C_r(X,Y) = \|J(X,Y-1) - J(X+1,Y)\| + C_u(X,Y) \tag{3.4}$$

Where C_l, C_r and C_u are image gradients resulting from nonadjacent pixels becoming neighbors when a seam pixel separating them is removed. This calculates forward energy superior to others as both color and intensity data is considered. The results of color segmentation and saliency map are given in Figure 3.2. It can be observed from the results given in Figure 3.2 that saliency map suppresses the noise effect and enhances the hand region.



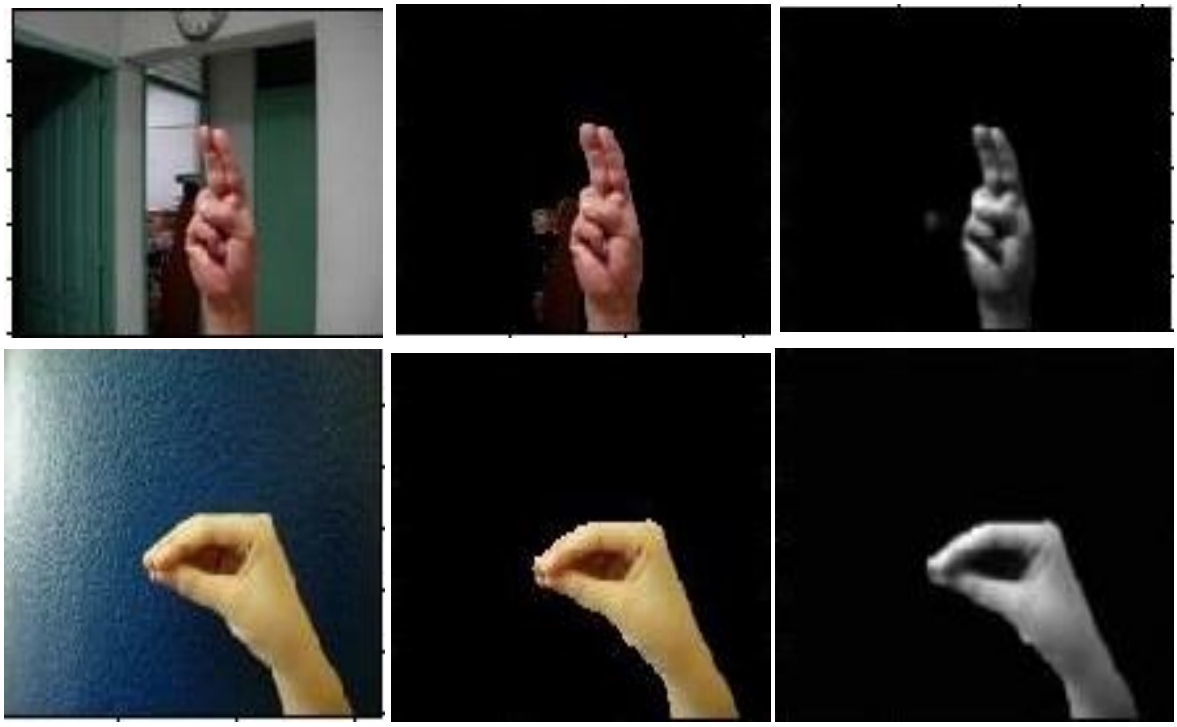


Figure 3. 2: Depiction of different hand postures, Column 1: Input hand postures of different datasets, Column 2: Skin likelihood of the image, Column 3: Saliency map of skin likelihood images.

3.2 Texture and Shape Feature Extraction

Good features are definitely needed for better classification. Features characterize a particular object in a well-defined way. As hand has all-out flexibility, even a slight change in the finger alignment of a particular posture needs attention. That's why to classify hand postures correctly more than one feature is required. Here two familiar feature extraction methods are used on hand image to get the texture and shape features. The Gabor filters [18] are used to extract texture features and Pyramid Histograms of Oriented Gradients (PHOG) [19] is used to extract shape features. The details of the two methods are described in the following sections.

3.2.1 The Gabor filter:

Gabor filters can capture the most significant visual properties such as orientation selectivity, spatial locality, and spatial frequency characteristics [18]. Considering the preferable properties, we chose the Gabor features to represent the hand gesture images. In the proposed method, we first convolved the segmented hand gesture images with the Gabor filters. Then we extract the feature vector by calculating the mean and variance of the Gabor filtered images. A 2-D Gabor filter is an oriented sinusoidal grating modulated by a 2-D Gaussian function, with a modulation frequency 'W', and is given in Eq. (3.5).

$$\check{\mathrm{G}}(x,y) = \check{\mathrm{g}}_{\sigma}(x,y)\exp(2\pi j\hat{\mathrm{W}}(xcos\theta + ysin\theta)) \tag{3.5}$$

Where,

$$\check{\mathrm{g}}_{\sigma}(x,y) = \left(\frac{1}{\sqrt{2\pi\sigma_x\sigma_y}}\right) exp\left(\frac{-1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right) \tag{3.6}$$

The Gabor filtered output of an image ∫(x, y) is obtained by the convolution of the image with the Gabor function Ğ(x, y). The parameters of a Gabor filter are the modulation frequency Ŵ, the orientation parameter θ and the scale σ of the Gaussian function. Local orientations and spatial frequencies explicit in Gabor filters are therefore used as the key features for texture processing. The input image is generally filtered by a family of Gabor filters tuned to several resolutions and orientations.

Texture representation

Texture classification is very important in image analysis. Content based image retrieval, inspection of surfaces, object recognition by texture, document segmentation are few examples where texture classification plays a major role.

Classification of texture images, especially those with different orientation and scale changes, is a challenging and important problem in image analysis and classification. This section describes texture representation based on Gabor Filter. A set of Gabor wavelets of different scale and orientation is convolved with an image to estimate the magnitude of local frequencies of that approximate scale and orientation. After applying Gabor filters on the image with different orientation at different scale, the energy content is calculated using Equation (3.7).

$$\bar{E}(r,s) = \sum_x \sum_y \left| \check{G}_{rs}(x,y) \right| \tag{3.7}$$

The mean μ_{rs} and standard deviation σ_{rs} of all transformed coefficients are found using Equations (3.8) and (3.9) respectively. These values represent the feature of the homogeneous texture image.

$$\mu_{rs} = \frac{\bar{E}(r,s)}{mn} \tag{3.8}$$

$$\sigma_{rs} = \sqrt{\frac{\bar{E}(r,s) - \mu_{rs}}{mn}} \tag{3.9}$$

A feature vector 'Ƒ' for texture representation is created using the mean and standard deviation as feature components. If R scales and S orientations are considered in the implementation, then the corresponding feature vector is given in Equation (3.10).

$$Ƒ = (\mu_{00}, \sigma_{00}, \mu_{01}, \sigma_{01}, \ldots\ldots \mu_{r-1\,s-1}, \sigma_{r-1\,s-1}) \tag{3.10}$$

By selectively changing each of the parameters of the Gabor filter, one can tune the filter to a specific pattern arising in the image. The Gabor filter response for 5 scale and 8 orientations is shown in Figure 3.3.

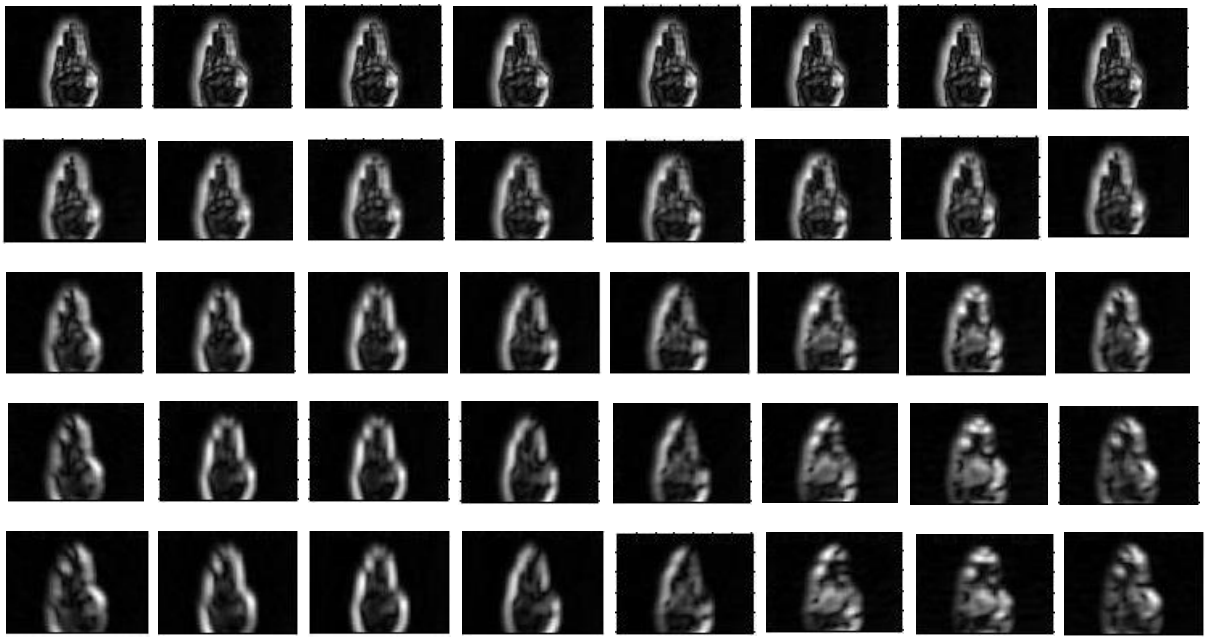


Figure 3. 3: Response of Gabor filter with 5 scale and 8 orientations.

3.2.2 Pyramidal Histograms of Oriented Gradients:

PHOG is basically a shape descriptor and it is mainly derived from two sources the image pyramid representation and the Histogram of Gradient Orientation [19] thus we can say that Pyramid HOG descriptor is a spatial pyramid representation of HOG descriptor. Here PHOG is used to extract the shape features of segmented hand region. In PHOG, first edges are extracted using the Canny edge detector. Then the image is divided into a series of increasing finer spatial grids by continually doubling the number of divisions in each axis direction. Along each dimension at resolution level $L = l$ the grid has 2^l cells. After division the orientation gradients are calculated using a 3×3 Sobel mask and Gaussian smoothing is omitted as omission of smoothing performed better in practice [25]. Histogram of edge orientations within each cell is divided into N bins, and histograms of the same level are concatenated into one sequence. PHOG descriptor used in this method is calculated using $L = 2$ levels, $N = 8$ bins and

range of [0-360]. PHOG method has better performance because the spatial information of local shapes is enhanced [26]. PHOG can be used to classification of smile expression [27] and secure image retrieval [28]. Shape spatial pyramid representation for an image and grids for levels l = 0 to l = 2 are shown in Figure 3.4 along with histogram representations corresponding to each level.

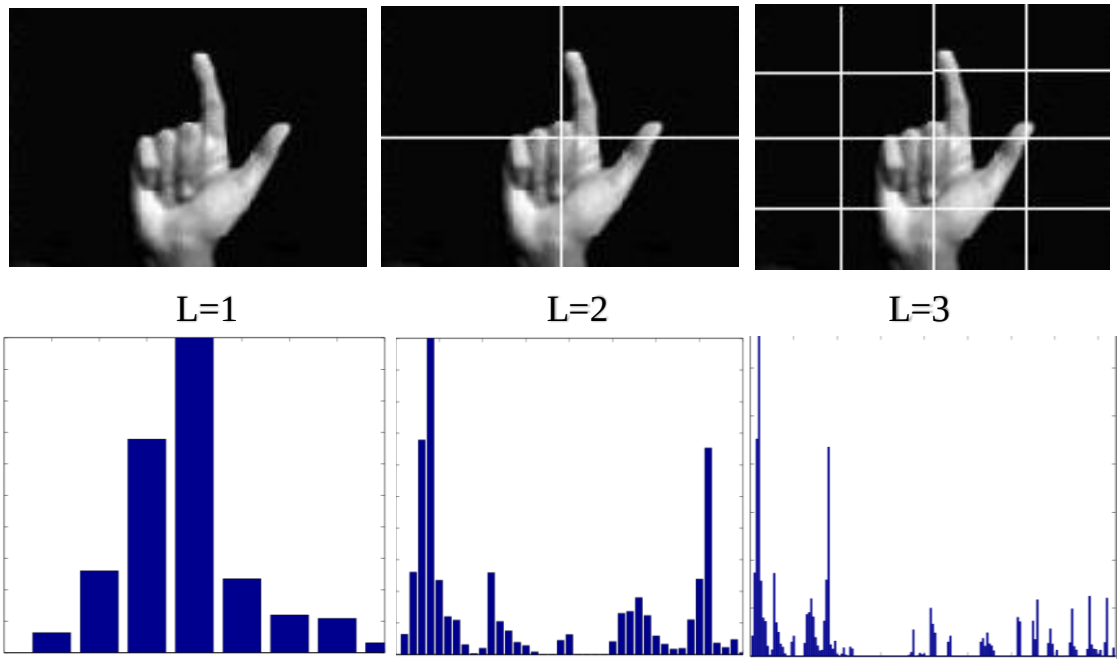


Figure 3. 4: Shape spatial pyramid representation. Top row: an image and grids for levels l = 0 to l = 2; below: histogram representations corresponding to each level.

From the Figure 3.4 it can be observed that, as we increase the level of computation of PHOG histogram, the magnitude of the histogram decreases but the details get finer and the representation improves.

CHAPTER-4

Experimental Result and Discussion

To evaluate the performance of proposed approach, an experiment is conducted using standard datasets i.e. Cambridge Hand Gesture Dataset [29], NUS hand posture datasets-I [30] and NUS hand posture datasets-II [31] and the average recognition rate (ARR) for each dataset is computed. The highest ARR achieved on these datasets is compared with the similar state of the art techniques. ARR is defined in equation (4.1) as:

$$ARR = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \text{ (In Percentage)} \quad (4.1)$$

Where TP, TN, FP and FN are the number of true positive, true negative, false positive, and false negative, respectively.

The sample images from NUS hand posture datasets-I are given in Figure 4.1. It contains 10 classes, 24 images for each class. In NUS hand posture datasets-I we have used 20 images for training and 4 images for testing which is in 5:1 ratio greater than the standard 10:1 ratio between training and testing images. The inter class difference between the postures of different classes is very minute which in turn makes the recognition task difficult. The 10 classes for NUS hand posture datasets-I are represented as $N1, N2, N3 \ldots\ldots\ldots N9, N10$ as shown in figure 4.1.

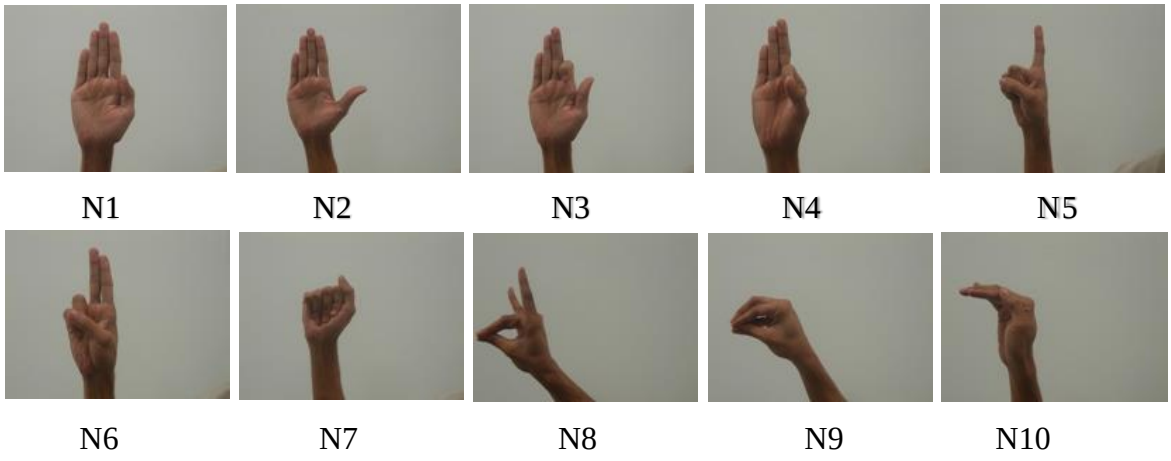


Figure 4.1: Sample images of NUS hand posture datasets I for 10 classes.

To get a good recognition rate the segmentation method and the extracted features needs to be excellent so that the difference within the classes would be identified. The recognition rate obtained for the NUS hand posture datasets-I is 97.5 % greater than the method described in [30]. The reason for this recognition rate is using YCbCr color space that has segmented the hand region effectively from the background as the background is of non-skin color or we can say background is quite simple and then the saliency map enhances the hand region. The features extracted using Gabor filter and PHOG are good enough to classify the interclass difference among hand postures using SVM Classifier efficiently.

The confusion matrix for the NUS hand posture datasets I is given in Table-1 and the Comparison of ARR with the other existing techniques for NUS hand posture datasets I is given in Table-2.

Table 1. Confusion Matrix for the Recognition Results of NUS hand posture datasets I

Predicted Classes

Input Classes	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10
N1	100	0	0	0	0	0	0	0	0	0
N2	0	100	0	0	0	0	0	0	0	0
N3	0	0	100	0	0	0	0	0	0	0
N4	0	0	0	100	0	0	0	0	0	0
N5	0	0	0	0	75	25	0	0	0	0
N6	0	0	0	0	0	100	0	0	0	0
N7	0	0	0	0	0	0	100	0	0	0
N8	0	0	0	0	0	0	0	100	0	0
N9	0	0	0	0	0	0	0	0	100	0
N10	0	0	0	0	0	0	0	0	0	100

Table 2. Comparison of ARR with the techniques of others for NUS hand posture datasets I

Method	Features	Classifier	ARR
Pisharady & Vadakkepat [30]	fuzzy-rough sets	SVM	94.5%
Our Method	Saliency Map & Gabor + PHOG	SVM	97.5%

In the modified, Cambridge Hand Gesture Dataset each class contains sequence of 57 images from which 50 images are used for training and 7 images are used for testing. The sample images from the dataset are given in figure 4.2. The recognition result obtained on this dataset is 98.752% which proves that the results obtained is independent from the varying light and illumination condition

proposed in the dataset. The Comparison of ARR with the other existing techniques for Cambridge Hand Gesture Dataset is given in Table-3.

The confusion matrix for the Cambridge Hand Gesture Dataset is given in Table-3 and the Comparison of ARR with the other existing techniques for Cambridge Hand Gesture Dataset is given in Table-4. The 9 classes of Cambridge Hand Gesture Dataset are represented as $C1, C2, C3 \ldots \ldots \ldots \ldots C8, C9$ as shown in figure 4.2.

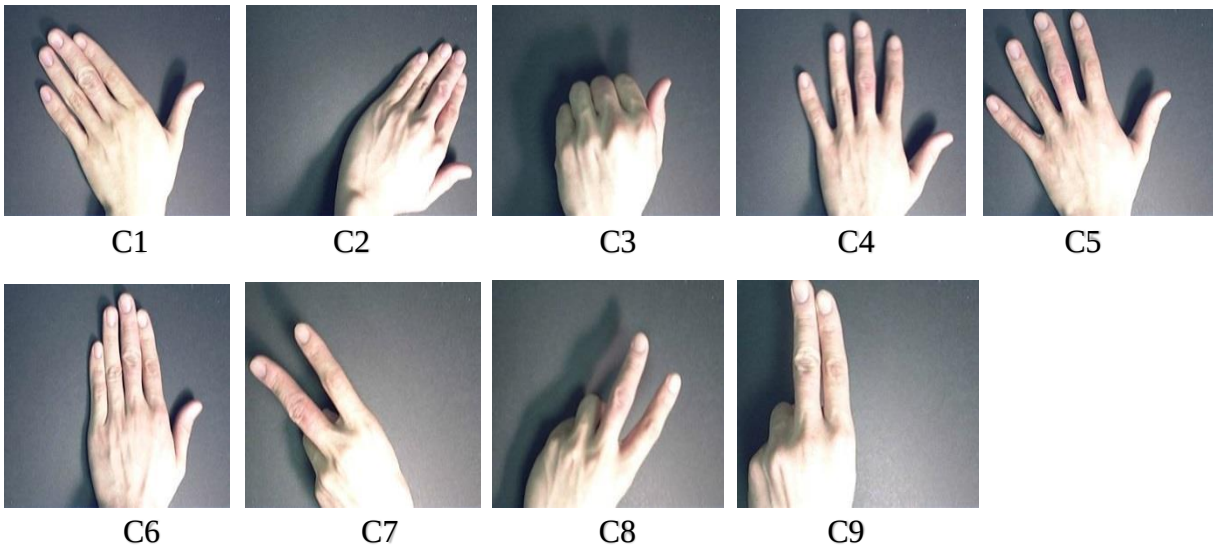


Figure 4.2: Sample images of Cambridge Hand Gesture Dataset for 9 classes

As we can see from the sample images that, in Cambridge Hand Gesture Dataset 9 postures are there, but if we look closely we would find that it consist of 3 basic hand shapes and 3 basic motions. Hence, we need to classify different shapes along with different motions at same time.

Table 3. Confusion Matrix for the Recognition Results of Cambridge Hand Gesture Dataset

Predicted Classes

Input Classes

	C1	C2	C3	C4	C5	C6	C7	C8	C9
C1	100	0	0	0	0	0	0	0	0
C2	0	100	0	0	0	0	0	0	0
C3	0	0	100	0	0	0	0	0	0
C4	0	0	0	100	0	0	0	0	0
C5	0	0	0	0	100	0	0	0	0
C6	0	0	0	0	0	100	0	0	0
C7	0	0	0	14.28	0	0	85.72	0	0
C8	0	0	0	0	0	0	0	100	0
C9	0	0	0	0	0	0	0	0	100

Table 4. Comparison of ARR with the techniques of others for Cambridge Hand Gesture Dataset

Method	Features	Classifier	Test scheme	ARR
Liu & Shao [32]	Spatio-temporal descriptors	SVM	LOO	85%
Gamal et al. [33]	Fourier Descriptor	SVM	LOO	98.5%
Baek et al. [34]	Local binary pattern	SVM	One-Against-AII(OAA)	97.33%
Our Method	Gabor + PHOG	SVM	Leave One Out(LOO)	98.572%

The sample images of NUS hand posture dataset-II are given in figure 4.3. It contains 10 classes, with 110 images for each class. In which 100 images are being used for training and 10 images for testing. Due to complex background the segmentation of hand gesture becomes difficult which in turn affects the recognition rate. From the above explained method we have secured a recognition rate of 94%. The use of YCbCr color segmentation limits the recognition rate here. Though saliency map to some extent is able to overcome the limitation imposed by YCbCr color segmentation, but if the background consist of skin color in larger part or whole background is of skin color then saliency map won't be able to generate the hand posture. The ARR achieved on NUS hand posture datasets-II justifies that Gabor filter gives rotation invariant and scale invariant features as this dataset contains postures on various scale and orientation.

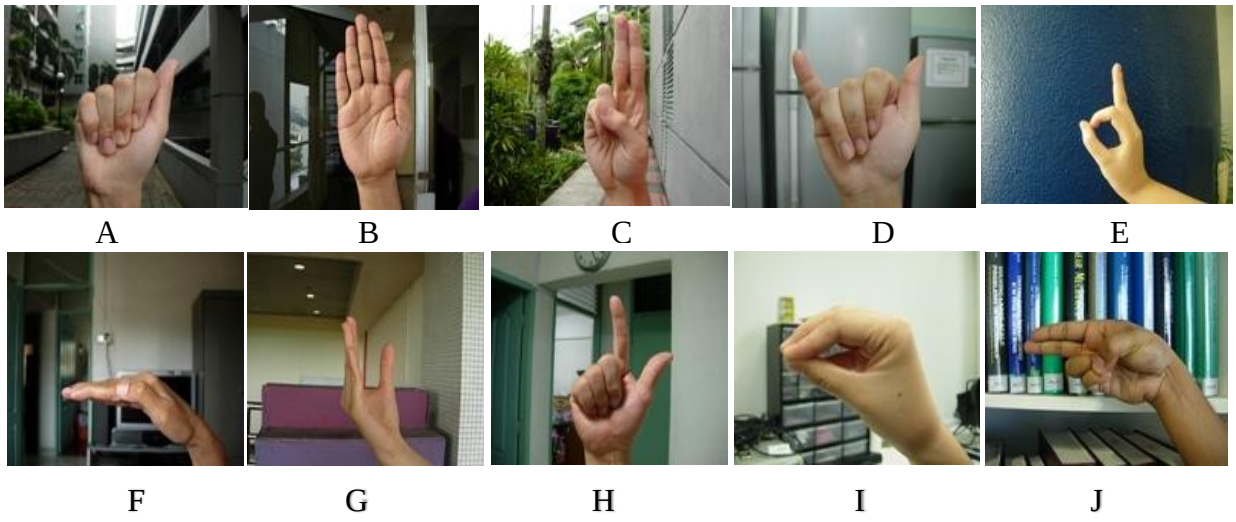


Figure 4.3: Sample images of NUS hand posture datasets II for 10 classes.

The confusion matrix for the NUS hand posture datasets II is given in Table-5 and the Comparison of ARR with the other existing techniques for NUS hand posture datasets II is given in Table-6.

Table 5. Confusion Matrix for the Recognition Results of NUS hand posture datasets II

Predicted Classes

Input Classes

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
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
E	0	0	0	0	100	0	0	0	0	0
F	20	0	0	0	0	80	0	0	0	0
G	10	0	0	0	0	0	90	0	0	0
H	0	0	0	0	0	0	0	100	0	0
I	0	0	0	0	0	0	0	0	100	0
J	0	10	0	0	0	20	0	0	0	70

Table 6. Comparison of ARR with the techniques of others for NUS hand posture datasets II

Method	**Features**	**Classifier**	**Test scheme**	**ARR**
Pisharady et al. [31]	Shape & Texture	SVM	LOO	94.36%
Chuang et al. [35]	Oconaire's skin model + image saliency	SVM	LOO	95.27%
Our Method	Gabor + PHOG	SVM	LOO	94%

CHAPTER-5

Conclusion and Future Scope

In this work, a novel approach for the recognition of hand posture is presented. The Skin-Saliency segmentation technique works quite well on a wide range of images. We have applied these segmentation technique on three standard data sets namely: NUS hand gesture dataset I, NUS hand gesture dataset II and Cambridge dataset. The texture and shape features are extracted using Gabor wavelet and pyramidal histogram of orientated gradient respectively. The hand poses are classified using a linear SVM classifier which uses the extracted features.

The performance of the proposed method is evaluated on three standard datasets and the recognition rate proves that it is robust and unfailing though the performance is limited to some extent if the complex background consist skin color on large section. The results are independent from varying light and illumination condition as well as rotation and scale invariant.

In future, the following extension can be done using proposed framework.

- Alteration in the proposed method to improve its recognition rate for images having skin color in background using some adaptive skin color model [36] [37] approach.
- Applying the proposed method on many other hand gesture dataset and checking the robustness. Though one size can never fit all, thus alterations in the method would be needed according to the characteristic or features of dataset.

- Until now we have worked only on static images. The work can be extended to evolve a real time hand gesture recognition system. That can be used to control computer [38] and mobile phones.
- Currently to control the TV we need to be dependent on remote control. Smart interactive television [39] [1] can be way forward using this method.
- Automobile drivers have been depended on hand postures to maneuver through traffic in past. New technology is facilitating automobile manufacturers to incorporate gesture recognition features in cars to let drivers control their systems of the cars. For example, a hand posture can initiate the in-car infotainment system, or other hand gesture can switch on the indicator. Driver's distraction have been a major source of concern to keep up safety on the road. Many times doing other things while driving people meet an accident. [40]Gesture centered car controlling systems would facilitate drivers to various things without even eyeing at the dashboard. Hyundai will be incorporating a "3-D gesture" control system in their top-end cars that will let a driver to control the audio system with just a wave of the hand.
- Gesture controlled wheelchair [41] [42] [43] for paraplegic patients could be a way forward in HGR applications.
- HGR for Human-Robot Interaction for service robot [44].

References

[1] D. Vishwakarma and R. Kapoor, "An Efficient Interpretation of Hand Gestures to Control Smart Interactive Television," *International Journal of Computational Vision and Robotics*, July, 2015.

[2] R. C. Gonzalez and R. E. Woods, Digital Image Processing.

[3] K. K. Rahini and S. S. Sudha, "Review of Image Segmentation Techniques: A Survey," *International Journal of Advanced Research in Computer Science and Software Engineering,* vol. 4, 2014.

[4] M. W. Khan, "A Survey: Image Segmentation Techniques," *International Journal of Future Computer and Communication,* vol. 3, 2014.

[5] D. K. Vishwakarma , R. Maheswari and R. Kapoor, "An Efficient Approach for the Recognition of Hand Gestures from Very Low Resolution Images," in *Fifth International Conference on Communication Systems and Network Technologies (CSNT)*, 2015.

[6] D. K. Vishwakarma and R. Kapoor, "Simple and intelligent system to recognize the expression of speech-disabled person," in *4th International Conference on Intelligent Human Computer Interaction (IHCI)*, 2012.

[7] G. Paschos, "Perceptually uniform color spaces for color texture analysis: an empirical evaluation," *IEEE Transactions on Image Processing,* pp. 932 - 937, 2001.

[8] I. T. Jolliffe, Principal Component Analysis, Springer-Verlag New York, 2002.

[9] S. Raschka, "Linear Discriminant Analysis," 2014.

[10] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Human Genetics*, vol. 7, no. 2, pp. 179-188, 1936.

[11] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988 - 999, 1999.

[12] B. Aisen, "A Comparison of Multiclass SVM Methods," http://courses.media.mit.edu/2006fall/mas622j/Projects/aisen-project/, 2006.

[13] Z. Ren, . J. Yuan, J. Meng and Z. Zhang, "Robust Part-Based Hand Gesture Recognition Using Kinect Sensor," *IEEE Transactions on Multimedia*, vol. 15, pp. 1110-1120, 2013.

[14] N. H. Dardas and N. D. Georganas, "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 11, pp. 3592 - 3607, 2011.

[15] S. Shiravandi, M. Rahmati and F. mahmoudi, "Hand gestures recognition using dynamic Bayesian networks," in *2013 3rd Joint Conference of AI & Robotics and 5th RoboCup Iran Open International Symposium (RIOS)*, Tehran, 2013.

[16] A. R. Várkonyi-Kóczy and B. Tusor, "Human–Computer Interaction for Smart Environment Applications Using Fuzzy Hand Posture and Gesture Models," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 5, pp. 1505 - 1514, 2011.

[17] Y. Gaus and F. Wong, "Hidden Markov Model-Based Gesture Recognition with Overlapping Hand-Head/Hand-Hand Estimated Using Kalman Filter," in *2012 Third International Conference on Intelligent Systems Modelling and Simulation*, 2012.

[18] M. Panwar, "Hand gesture recognition based on shape parameters," in *International Conference on Computing, Communication and Applications*, 2012.

[19] D. K. Ghosh and S. Ari, "A static hand gesture recognition algorithm using k-mean based radial basis function neural network," in *8th International Conference on Information, Communications and Signal Processing (ICICS)*, 2011.

[20] Arivazhagan, Ganesan and Priyal, "Texture classification using Gabor wavelets based rotation invariant features," *elsevier*, 2006.

[21] A. Bosch, A. Zisserman and X. Munoz, "Representing shape with a spatial pyramid kernel," in *CIVR'07*, 2007.

[22] D. Chai and K N. Ngan, "Face Segmentation Using Skin Color Map in Videophone Applications," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, 1999.

[23] S. L. Phung, A. Bouzerdoum and D. Chai, "Skin Segmentation Using Color Pixel Classification: Analysis and Comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, 2005.

[24] R. Achanta and S. Susstrunk, "Saliency Detection for Content-Aware Image Resizing," in *16th IEEE International Conference on Image Processing (ICIP)*, 2009.

[25] S. Avidan and A. Shamir, "Seam carving for contentaware image resizing," *ACM Transactions on Graphics*, vol. 26, 2007.

[26] Y.-S. Wang, C.-L. Tai, O. Sorkin and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Transactions on Graphics*, vol. 27, 2008.

[27] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Conference on Computer Vision and Pattern Recognition*, 2005.

[28] Z. Li, J.-i. Imai and M. Kaneko, "Facial-component-based Bag of Words and PHOG Descriptor for Facial Expression Recognition," in *IEEE International Conference on Systems, Man, and Cybernetics*, 2009.

[29] Jun Chen and Yang Bai, "Classification of Smile Expression using Hybrid PHOG and Gabor features," in *2010 International Conference on Computer Application and System Modeling*, 2010.

[30] M.Madlin Asha and Dr.J.Jennifer Ranjani, "Secure Image Retrieval using PHOG Descriptor," in *International Conference on Advanced Computing and Communication Systems*, 2013.

[31] T. Kim and R. Cipolla, "Canonical Correlation Analysis of Video Volume Tensors for Action Categorization and Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI),* vol. 31, no. 8, pp. 1415-1428, 2009.

[32] P. K. Pisharady and P. Vadakkepat, "Hand Posture and Face Recognition Using a Fuzzy Rough Approach," *International Journal of Humanoid Robotics*, vol. 39, no. 2, pp. 23-42, 2010.

[33] P. K. Pisharady, P. Vadakkepat and . A. P. Loh, "Attention Based Detection and Recognition of Hand Postures Against Complex Backgrounds," *Springer*, 2012.

[34] L. Liu and L. Shao, "Synthesis of Spatio-Temporal Descriptors for Dynamic Hand Gesture Recognition Using Genetic Programming," in *0th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 2013.

[35] H. M. Gamal, H. Abdul-Kader and E. A. Sallam, "Hand Gesture Recognition using Fourier Descriptors," in *17th International Conference on Pattern Recognition*, 2004.

[36] J. Baek, J. Kim and E. Kim, "Comparison Study of different Feature Classifiers for Hand Posture Classification," in *13th International Conference on Control, Automation and Systems*, 2013.

[37] Y. Chuang, L. Chen and G. Chen, "Saliency-guided improvement for hand posture detection and recognition," *Neurocomputing*, pp. 404-415, 2014.

[38] M. Soriano, B. Martinkauppi, S. Huovinen and M. Laaksonen, "Adaptive skin color modeling using the skin locus for selecting training pixels," *Pattern Recognition,* vol. 36, no. 3, p. 681–690, 2003.

[39] A. Y. Dawod, J. Abdullah and M. J. Alam, "Adaptive skin color model for hand segmentation," in *2010 International Conference on Computer Applications and Industrial Electronics (ICCAIE)*, 2010.

[40] S. Wan and H. T. Nguyen, "Human computer interaction using hand gesture," in *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008.

[41] S.-H. Lee, M.-K. Sohn, D.-J. Kim and B. Kim , "Smart TV interaction system using face and hand gesture recognition," in *2013 IEEE International Conference on Consumer Electronics (ICCE)*, 2013.

[42] X. H. Wu, M.-C. Su and P.-C. Wang, "A Hand-Gesture-Based Control Interface for a Car-Robot," in *The 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010.

[43] S. P. Kang , G. Rodnay, M. Tordon and J. Katupitiya, in *2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2003.

[44] L. A. Rivera and G. N. DeSouza, "A power wheelchair controlled using hand gestures, a single sEMG sensor, and guided under-determined source signal separation," in *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 2012.

[45] R. Posada-Gomez, L. H. Sanchez-Medel , G. A. Hernandez and A. Martinez-Sibaja , "A Hands Gesture System Of Control For An Intelligent Wheelchair," in *4th International Conference on Electrical and Electronics Engineering ICEEE 2007.*, 2007.

[46] R. C. Luo and . Y.-. C. Wu, "Hand gesture recognition for Human-Robot Interaction for service robot," in *2012 IEEE Conference onMultisensor Fusion and Integration for Intelligent Systems (MFI)*, 2012.