# SIGN LANGUAGE INTERPRETATION USING HAND GESTURE RECOGNITION

**By**

**Monika Lodha**
**(05MCE007)**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**INSTITUTE OF TECHNOLOGY**

**NIRMA UNIVERSITY OF SCIENCE & TECHNOLOGY**

**Ahmedabad 382481**

**May 2007**

# SIGN LANGUAGE INTERPRETATION USING HAND GESTURE RECOGNITION

**Major Project**

Submitted in partial fulfillment of the requirements

For the degree of

**Master of Technology in Computer Engineering**

**By**

**Monika Lodha**
**(05MCE007)**

Guide
**Prof. S.N. Pradhan**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**INSTITUTE OF TECHNOLOGY**

**NIRMA UNIVERSITY OF SCIENCE & TECHNOLOGY**

**Ahmedabad 382481**

**May 2007**

This is to certify that Dissertation entitled

# SIGN LANGUAGE INTERPRETATION USING HAND GESTURE RECOGNITION

Presented by

Monika Lodha.

has been accepted toward fulfillment of the requirement

for the degree of

Master of technology in Computer Science & Engineering

| | |
|---|---|
| Prof. S. N. Pradhan | Prof. D. J. Patel |
| Professor In Charge | Head of The Department |

Prof. A. B. Patel

Director, Institute of Technology

# CERTIFICATE

This is to certify that the Major Project entitled "SIGN LANGUAGE INTERPRETATION USING HAND GESTURE RECOGNITION" submitted by Ms. Monika Lodha (05MCE007), towards the partial fulfillment of the requirements for the degree of Master of Technology in Compute Science & Engineering of Nirma University of Science and Technology, Ahmedabad is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project, to the best of my knowledge, haven't been submitted to any other university or institution for award of any Master degree.

Project Guide

Dr. (Prof.) S. N. Pradhan

P. G. Coordinator
Department of Computer Engineering,

Institute of Technology,

Nirma University,

Ahmedabad

Co. Guide

Prof. M. D. Bhavshar

Department of Computer Engineering,

Institute of Technology,

Nirma University,

Ahmedabad

Date :-

ACKNOWLEDGEMENT

It gives me immense pleasure in expressing thanks and profound gratitude to **Dr**. **S.N Pradhan**, Professor, and **Smt. Madhuri Bhavsar,** Asst professor, Computer Science Department, Institute of Technology, Nirma University, Ahmedabad for their valuable guidance and continual encouragement throughout the Major project . I am heartily thankful to them for their precious time, suggestions and sorting out the difficulties of my topic that helped me a lot during this study.

I would like to give my special thanks to Prof. **D.J Patel**, Head, Department of Computer Engineering, Institute of Technology, Nirma University, Ahmedabad for his encouragement and motivation throughout the Major Project , I am thankful to **Mr. B. Gopalkrishnan**, Scientist, ISRO, Ahmedabad, I am also thankful to **Prof. A. B. Patel**, Director, and Institute of Technology for his kind support in all respect during my study.

I am thankful to all faculty members of Computer Engineering Department, Nirma University, and Ahmedabad for their special attention and suggestions towards the project work.

The blessings of God and my family members makes the way for completion of major project-I. I am very much grateful to them.

I am thankful to all those friends who always support and encourage me.

**Monika Lodha**
Roll No. 05MCE007

# ABSTRACT

Hand gesture recognition is an active area of research in computer vision with wide range of application in the area of sign language recognition, games, human computer interaction etc. Sign language is the basic communication method with speech and hearing imapaired. The aim of this thesis work is to develop a system which will translate hand gesture into the corresponding sing language entity. The various stages of the system are; segmentation, feature extraction and classification.

.

The Segmentation method is used to locate hands in the images. . Gaussian model is used to find the skin-color distribution. This is further used to obtain likelihood of skin for any pixel of image. These images are transformed into binary image by proper thresholding. The binary image of the hand gesture is processed with erosion to eliminate the noise and then processed with dilation to recover the original shape.

Three methods for recognition and classification of hand gesture are presented-Boundary contour method, Medial axis (Skeletonization) method and Circle based method. Common gesture of fingers to indicate the numerals have been used for testing. The results indicate that it is possible to identify the number from the gesture of fingers.

# CONTENTS

# List of Figures

# List of Tables

# Nomenclature

| | |
|---|---|
| COM | Center Of Mass |
| COG | Center Of Gravity |
| DOF | Degree Of Freedom |
| RGB | Red Green Blue |

# 1 INTRODUCTION

## 1.1 GENERAL

Video based hand gesture recognition is an active research area occupying a range of disciplines from practical sign language study to highly theoretical image processing. Humans are highly literate in gestural communication. Every interaction with the physical world involves some form of physical manipulation which may be considered as gesture. Conversation between individuals involves not only speech, but also a rich vocabulary of gesture which under constraints of noise or language may be the primary mode of communication [1]. One application of gesture recognition is sign language interpretation. Sign language is useful for hearing and speech impaired people.

A gesture language can be rich enough to convey all the meaning of spoken and written words, as the various sign languages of the world demonstrate.

## 1.2 MOTIVATION FOR HAND GESTURE RECOGNITION

Sign language is one form of communication for hearing and speech impaired [2]. Sign languages are expressed by the gesture of hand or head. Hand gesture is most effective and intuitive method used for everyday human communication. Gesture recognition is the interpretation of a given gesture into text form. There are two types of gestures namely static gestures and dynamic gestures. Static gesture recognition determines the action depending on a gesture in an image. Dynamic gesture recognition determines the action according to the hand motion. Each dynamic gesture is represented by a group of image frames. In this approach, gestures are taken under the constraint of uniform background. The focus of this project is on static gesture with a single hand. Here hand gesture classification is done basically for number from 1-5.

## 1.3 SYSTEM CHALLENGES

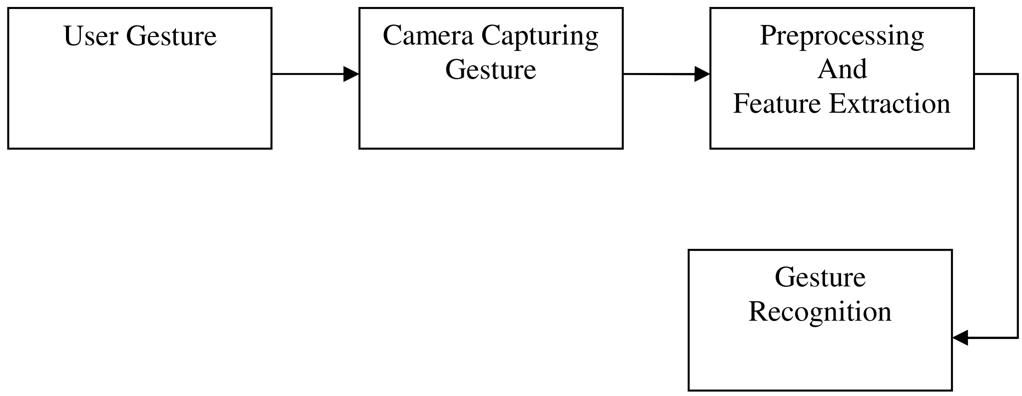



Figure 1.1 Block Diagram of hand gesture recognition system

Figure 1.1 shows the steps involved in the gesture recognition system. Each step is an area of research in its own right.

Human hand has a large number of degree of freedom modeling structural and dynamic feature for gesture recognition of such a complex object is a tough task. Simplification of the model reduces the number of separable gestures and may impose constraints on the user for accurate operation. An example is the use of fingertip positions and palm position to characterize gesture which depends on the user holding the palm rigid for reliable recognition.

Implementing real life gesture recognition system is a daunting task considering the challenges at each step of the system like background clutter, dynamic illumination changes, camera movements etc. The performance of recognition steps depends on the previous stages and also on the choice of feature for gesture representation.

## 1.4 Problem Definition

This Project is mainly concerned with the implementation of visual gesture recognition system, which replaces use of electro-mechanical devices and sensors for a sign language interpretation. Different approaches have been presented for recognizing gesture for number system. Fuzzy c-means clustering algorithm is used for recognizing different gesture. The static hand gesture recognition system involves image capturing, segmentation, morphological filtering, feature extraction and different approaches for classification like boundary contour, skeltonization, and circle construction techniques.

## 1.5 SCOPE OF WORK

The video based hand gesture recognition is a challenging field in the image processing research area. One of the main implementation of the gesture recognition techniques is applicable for sign language interpretation. There are two types of gestures namely static gestures and dynamic gestures. Static gesture recognition determines the action depending on a gesture in an image. Dynamic gesture recognition determines the action according to the hand motion. The focus of this project is on static gesture with a single hand. The main case study for implementation involves recognition of numbers. Other aspects to static gesture recognition is for alphabets, actions etc. But the main scope of this thesis work is limited to recognition of numbers represented by fingers of human hand.

## 1.6 ORGANIZATION

The Chapter 2 gives an overview of relevant literature and status research in this field. Theoretical background used in the system is discussed in Chapter 3. System outputs and results are presented in chapter 4 and finally conclusion is given in chapter 5.

# 2 LITERATURE SURVEY

## 2.1 GENERAL

Using the system block diagram given in the last chapter; the various research and commercial system related to various steps used in the system documented in the literature are discussed.

## 2.2 SEGMENTATION

Any video stream of gestures will contain a large proportion of redundant information which must be removed in order to reduce the amount of data for analysis by isolating the areas of interest. This is particularly the case for systems which allow dynamic location (SPDL or DPDL) gestures which must capture an area around the user sufficient for natural movement. The sheer volume of data contained in digital video makes this data reduction necessary but computationally intensive in itself.

The object of segmentation is dictated by the final feature space required. For most gesture systems it is either the hand or the hand and arm region. Some view based feature extraction techniques work directly on the input stream and may be considered to combine segmentation and extraction work directly on the input stream and may be considered to combine segmentation and extraction. A binary image representing areas of interest is the most common output from the segmentation stage [1].

### 2.2.1 Difference Based Segmentation

The motivation for difference-based segmentation schemes is to isolate regions of change within the input images as the areas of interest. In the most common system configuration using a fixed camera in a environment static apart from the gesturer, this naturally leads to the notion of the *background* which is the image of those static fixtures in the field of view and the *foreground* which is the moving gesturer.

The simplest form of difference-based segmentation use image of the static environment and subtracts this image from the current frame to establish regions of difference. This procedure, known as *background subtraction* is common in the literature ([1, 3]) and is effective for low computational cost. Each pixel in the current frame is simply subtracted from the background image and the magnitude of the difference is compared to a threshold. A foreground mask may be described according to Eq. (2 .1) in which *C* is the current frame, *B* is background image and *Tb* is the threshold. The threshold must be set above the noise level in the image but low enough to detect any valid image change.

$$F = \begin{Bmatrix} 1 \, if \;\; |C(x, y) - B(x, y)| > Tb \\ 0 \, otherwise \end{Bmatrix} \qquad \ldots (2.1)$$

Whilst it is computationally efficient and straightforward to implement, background subtraction suffers from a number of limitations. The environment must be static without areas of color (or gray level for monochrome imaging) similar to the foreground objects. A clear background image must also be obtained before segmentation can begin. Even when these constraints are met, the technique is not robust to changes in image illumination unless the image subtraction is done in some normalized color space.

### 2.2.2 Color Based Segmentation

Skin color can be used as an important clue for isolating area of interest from the input image. Complete color representation are also subject to variation from differing illumination condition. So the color space chosen should be such that it will be less sensitive to illumination variation.
A considerable obstacle with color-based techniques is that skin appears to vary greatly in color from one human to another. This may be overcome with initialization to the skin tone of a particular user or by capturing the overall distribution of human skin coloration [1].

#### *2.2.2.1 Colorspaces Used For Skin Modeling*

Colorimetry, computer graphics and video signal transmission standards have given birth to many color spaces with different properties. A wide variety of them have been applied to the problem of skin color modeling [1].

Three different color models are common in image processing: RGB, HSV and YIQ (also known as YUV). The name of each model is an acronym from the three components used to represent a color in each model. RGB and YCbCr are hardware derived models, whereas HSV is based upon human perception of color. Table 2.1 gives a description of each of the color models and its constituent components.

Table 2.1: Color Models

| Model | Components | | |
|---|---|---|---|
| RGB | Red | Blue | Green |
| HSV | Hue | Saturation | Value |
| YCbCr | Luminance | (blue-yellow) | (red-green) |

Most color-based segmentation schemes use the RGB color representation as it is the most common format for digital video input and display .These systems are subject to problems with changes in illumination and differing skin tones amongst users.

Invariance to illumination changes is attempted by using RGB colors normalized. Eq (2.2) in which the RGB component values are substituted in the obvious way. Where the denominator sums include a 1 to avoid division by zero and give ranges of [0*;* 1]. Note that the blue component effectively becomes redundant in this representation. Chromatic color (r, g) can be defined as

$$r = \frac{R}{R+G+B} \quad \text{... (2.2)}$$

$$g = \frac{G}{R+G+B} \quad \text{... (2.3)}$$

YCbCr color space can be defined as

$$Y = 0.2126R + 0.7152G + 0.0722B \quad \text{... (2.4)}$$

$$Cb = Y - B \quad \text{... (2.5)}$$

$$Cr = Y - R \quad \text{... (2.6)}$$

### *2.2.2.2 Histogram Based Technique*

The color histogram revealed that the distribution of skin-color of different people are clustered in the chromatic color space and a skin color distribution can be represented by a Gaussian model N(m, C), where:

Mean: $m = E\{x\}\,\text{where x} = (\text{cb cr})^{\text{T}}$ ... (2.7)

Covariance: $C = E\{(x\ ?\ \text{m})(\text{x?m})^{\text{T}}\}$ ... (2.8)

With this Gaussian fitted skin color model, we can now obtain the likelihood of skin for any pixel of an image. Therefore, if a pixel, having transform from RGB color space to chromatic color space, has a chromatic pair value of (Cb,Cr), the likelihood of skin for this pixel can then be computed as follows [4]:

$$\begin{gathered} Likelihood = \text{P(cb, cr)} = \exp[-.5(\text{x - m})^{T}\ \text{c - 1(x - m)}] \\ \text{where : x} = (\text{cb, cr})^{T} \end{gathered} \quad \text{... (2.9)}$$

### *2.2.2.3 Skin Segmentation*

Beginning with a color image, the first stage is to transform it to a skin-likelihood image. This involves transforming every pixel from RGB representation to chroma representation and determining the likelihood value based on the equation given in the previous section. The skin-likelihood image will be a grayscale image whose gray values represent the likelihood of the pixel belonging to skin.

Since the skin regions are brighter than the other parts of the images, the skin regions can be segmented from the rest of the image through a thresholding process. To process different images of different people with different skin, a fixed threshold value is not possible to be found. Since people with different skins have different likelihood, an *adaptive thresholding* process is required to achieve the optimal threshold value for each run.

The adaptive thresholding is based on the observation that stepping the threshold value down may intuitively increase the segmented region. However, the increase in segmented region will gradually decrease (as percentage of skin regions detected approaches 100%), but will increase sharply when the threshold value is considerably too small that other non-skin regions get included. The threshold value at which the minimum increase in region size is observed while stepping down the threshold value will be the optimal threshold [4].

### 2.2.3 Marker-Based Segmentation

A number of gesture recognition systems in the literature facilitate the process of Segmentation by using easily identified markers to locate relevant image areas.
A multi-coloured glove is used that allow hand to be segmented according to these known colours and to facilitate feature extraction by providing known regions for each hand feature.

Maggioni and K¨ammerer [18] presented a marker-based approach using a black and white marker consisting of eccentric circles on a glove. Segmentation is performed using simple binary thresholding and feature extraction is facilitated by exploiting the geometric properties of the marker [1].

## 2.3 Feature Extraction

Feature extraction may provide feature parameters to a subsequent classification stage, or may be the final output of a gesture recognition system if the feature space itself provides useful information. Analysis of feature parameters from the segmented image also allows feedback to the segmentation module to validate the selected areas of interest [1].

For gesture recognition systems that include a classification stage, the feature space must be chosen to allow separation of the output gesture classes.

### *2.3.1 Parameter Computation*

The process of combining image features and apriori constraints to generate feature vectors is called *constraint fusion*. These constraints may allow segmentation candidates to be validated for feature extraction or may be used to disambiguate or dictate image feature to model feature mappings. This is the problem of producing the feature vector which contains the model parameters from the segmentation output [1].

Various techniques are used to generate model parameters, but most use a successive approximation structure [1]. This structure may be seen in Fig. 2.1, where the comparison loop continues until the difference between image features and model features is suitably small.

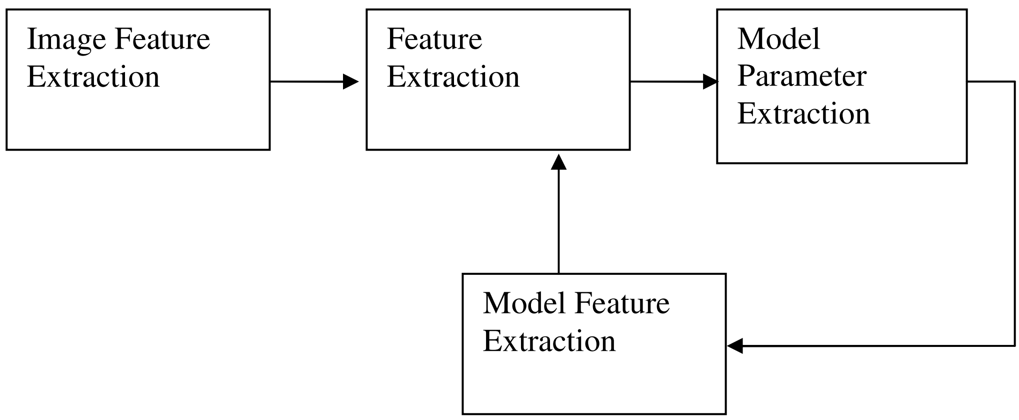



Figure 2.1 Model-based Parameter Computation

### *2.3.2 Orientation-Histogram Extraction*

It is require gestures to be the same regardless of where they occur with the images boarders. To achieve this we will ignore position altogether, and tabulate a histogram of how often each orientation element occurred in the image. Clearly, this throws out information and some distinct images will be confused by their orientation histograms. In practice, however, one can choose a set of training gestures with substantially different orientation histograms from each other.

One can calculate the local orientation using image gradients. The outputs of the x and y derivative operators will be dx and dy. Then the gradient direction is atan (dx, dy) [5]. This is an appearance-based approach as the image properties are used directly to derive the feature space without being matched to a model.

### *2.3.3 Contour-Based Extraction*

Contours are used in image processing as a way of extracting information about the shape of a region of interest. An image contour is a closed boundary defining the connected region it contains. Contours may be generated most simply from regions of constant pixel color but may easily be made using more advanced boundary definitions. Representation of the contour can contain a complete description of the extracted contour or may approximate it in some format that allows compression, feature extraction or classification [1].

Gesture recognition for number system using contour representing hand outline is described by Kenny Teng, Jeremy Ng, Shirlene Lim. They identify the point of local extreme curvature in the contour as feature and classify the based on the number and height of peaks [2].

### *2.3.4 Skeleton- Based Feature Extraction*

A skeleton is a compact representation of an object (hand region in our case). The desired properties of the skeleton are preserving the topology of the object (the same number of connected components and the same number of holes), robust against translation, rotation, and scaling, and thin (no more than 2 pixels in width). The extraction of the hand region skeleton is performed using a distance transformation-based method: the chamfer distance transformation. The distance transformation is computed to extract the median axis of the hand region (local maxima). Then the resulting set is connected in order to preserve the object topology [6].

To see how this works, imagine that the foreground regions in the input binary image are made of some uniform slow-burning material. Light fires simultaneously at all points along the boundary of this region and watch the fire move into the interior. At points where the fire traveling from two different boundaries meets itself, the fire will extinguish itself and the points at which this happens form the so called `quench line'. This line is the skeleton [2].

Gesture recognition for number system using skeleton representing hand connectivity is described by Kenny Teng, Jeremy Ng, Shirlene Lim. They identify the branch point, normal point and end point and the length of branch to classify the gesture representing number [2].

### *2.3.5 Circle construction and feature extraction*

In this method circle is constructed by taking radius as the .7 of the farthest distance from COG in such a way that it should intersect all the fingers active in the particular gesture.

Gesture recognition for number system based on circle drawing such that it should cover all the active finger of gesture is described by *Asanterabi Malima, Erol Özgür, and Müjdat Çetin*. They used 0-1 transition property to classify the gesture representing number [7].

## 2.4 Classification

Classification is the mapping from a feature state or a sequence of feature states to an output gesture. In general, gesture classification includes two significant problems: firstly, optimal partitioning of the feature or feature-time space and secondly, finding an efficient classification procedure. For static gesture recognition systems the problem of partitioning is limited to the feature space. Dynamic gesture recognition requires the inclusion of time in the partitioned space which makes the partitioning and classification significantly more complicated [1].

### 2.4.1 Static Methods

The partitioning of the feature space into classes requires that each class be established on the basis of a region or representative members in the feature space. Class representatives can be given explicitly or established through a supervised learning procedure from examples.

When the feature space is sufficiently small correlation classification can be used. Here classification is performed by correlating the feature vector with each class representative vector. The correlation result for each class representative is compared to a threshold determined offline by the correlation between that representative and the representatives of the other classes. They also present a general classifier for nearest-neighbor classification based on decision surfaces [1].

*2.4.1.1 Fuzzy-C-Means clustering algorithm*

During the training operation, feature vectors ($x1;x2;x3; : : : : ; xn$) are calculated for all gesture images. Where $x_k$ is the set of intensity of average intensity of each block which is obtain by dividing bounding box containing hand region presenting hand gesture. Next step is the classification of these feature vectors based on similar data by using clustering technique . Generally K-means clustering (crisp) or Fuzzy-C-means clustering (fuzzy) techniques are preferred as the clustering techniques for classification ofdata vectors. But K-means clustering algorithm is sensitive to initial randomly selected clustercenters and moreover it does not find the optimal solution. On the contrary, Fuzzy-C-Means algorithm is not sensitive to initial cluster centers [8].

### 2.4.2 Dynamic Methods

Dynamic gesture recognition introduces time into the feature space to be partitioned. Time partitioning is usually based upon a model of human gesture as consisting of triphasic cycles of *preparation*, *stroke* and *retraction* separated by rest states. Gesture classification is usually performed with a temporally invariant technique as the period spent in each gesture phase may vary without changing the intended gesture.

The temporal context of specific gestures within the sequence of gestures may also be used aid gesture classification. This approach is motivated by the ubiquity of grammar in any natural language which defines valid sequences of symbols, in this case gestures [1].

## 3.1 GENERAL

Hand gesture recognition system uses consumer-level computer hardware and video devices to capture the hand gestures of a user. These gestures are further processed for sign language interpretation.

## 3.2 HAND ANATOMY AND HAND MODEL

The anatomy of the hand defines the structures and range of movement that the hand may take. Hand models are derived from hand anatomy to capture this range of movement and the intrinsic constraints that the physical structure of the hand imposes. Many applications for hand models may be found in gesture recognition and in computer graphics animation, so the general problem is well researched. The particular problem of generating hand model parameters from images motivates hand models which derive the hand pose from easily identifiable features using the physical constraints of hand anatomy to disambiguate the complicated form of the hand.

Most hand models are derived from the skeletal structure of the hand. Fig3.1 gives a complete view of the human hand skeleton and its 27 bones. The bones are divided into three groups: *carpals* (wrist bones), *metacarpals* (palm bones) and *phalanges* (finger bones). Anatomical analysis of the hand yields that there re around 27 Degree of Freedom (DOF) within the human hand.

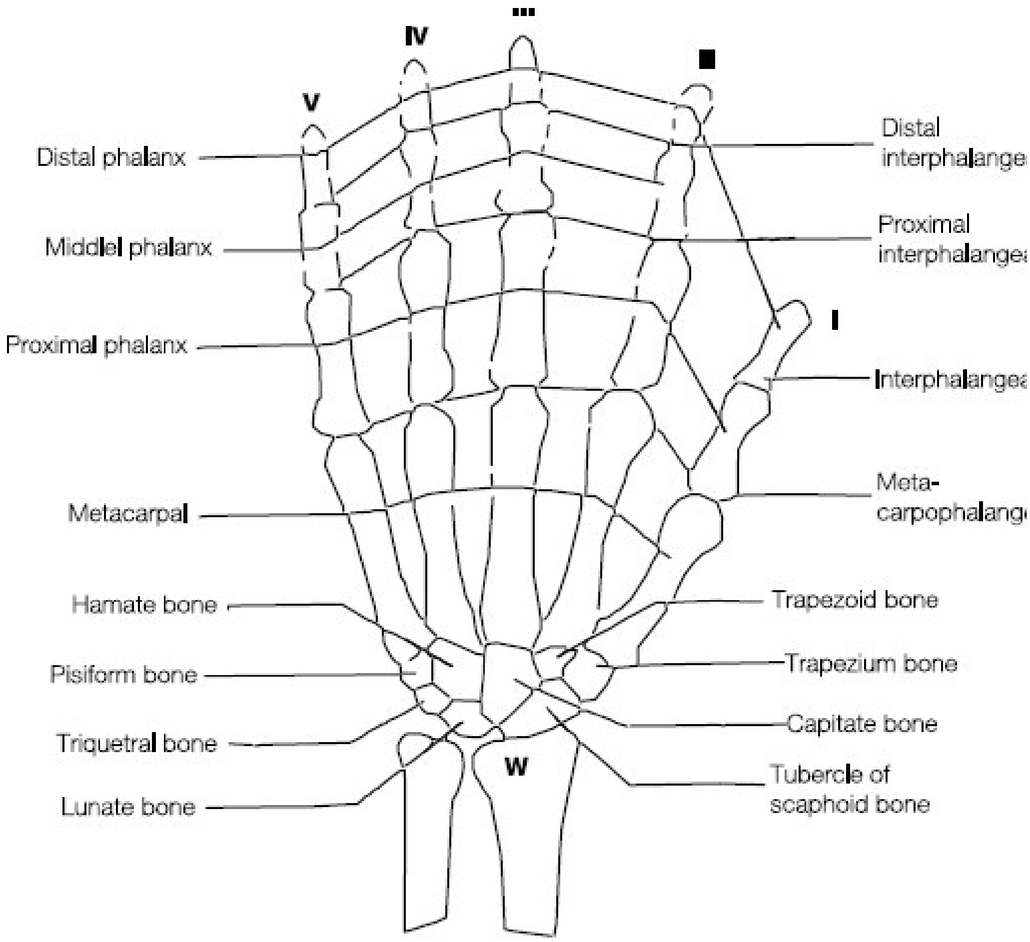



Figure 3.1 Hand Model

The small number of DOF in comparison to the number of bones in the hand implies that most joints have considerable limitations on their freedom of movement. For example, the interphangial (IP) joints in the fingers have only one degree of freedom (extension / flexion) and the metacarpalphalangeal (MCP) joints (knuckles) have two degrees of freedom as they can also move the finger from side-to-side (adduction / abduction). The extent of freedom within each degree is also limited: using the MCP joints for example again, the adduction / abduction degree is limited to movement through a range of around pi/6 and the extension / flexion degree to around pi/2.

Dependencies also exist between the degrees of freedom of single joints and of Different joints, further reducing the range of possible hand configurations. For instance, curling the fingers into the palm has been found to limit the ability of the finger to abduct or adduct.

## 3.3 STATIC HAND GESTURE RECOGNITION SYSTEM

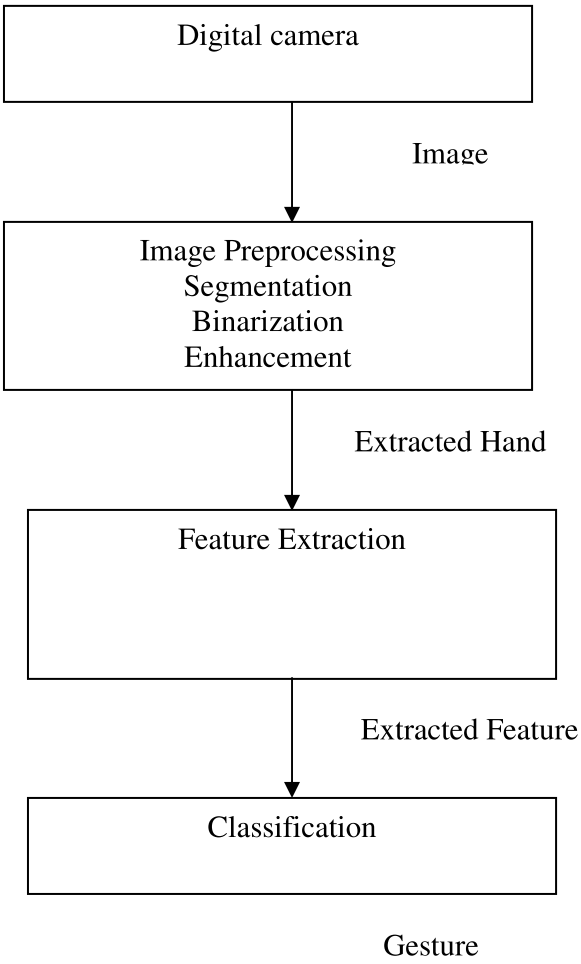



Figure 3.2 Static hand gesture recognition system

### 3.3.1 Preprocessing

Preprocessing of the image starts with image capturing, segmentation of the hand, morphological filtering, and feature extraction.

After capturing the image, the hand is segmented out from the rest of the image by using a color attribute. It is required that color space chosen should be such that it will be less sensitive to illumination variations. Color model chosen in this case is Ycbcr. A skin color distribution can be represented by a Gaussian model N(m,C), refer equation 2.4-2.8

With this Gaussian fitted skin color model, we can now obtain the likelihood of skin for any pixel of an image. Therefore, if a pixel, having transform from RGB color space to chromatic color space, has a chromatic pair value of (cb,cr), the likelihood of skin for this pixel can then be computed as give in equation 2.9.

After finding skin likelihood image by applying proper thresholding binary image is obtained. After thresholding, morphological filtering techniques like erosion and dilation operations are performed for noise elimination.

### 3.3.2 Feature Extraction

Different techniques have been used to analyze and classify the hand gestures, namely boundary contour, skeletonization and drawing circle such that it should contain all the finger presented in gesture. Based on this different method features are generated which are used for the classification of gesture.

### 3.3.3 Boundary Contour

Boundary contour is the process of determining the Euclidean distance of any point on the edge of an image to the center of mass (COM). The peaks (maximum) are the furthest point from the COM, that is, they represent the positions of the tip of the fingers. Proper thresholding is used to count the number of peaks to be considered as the tip of finger. Finally the number of peaks obtain gives the number presented by gesture.

### 3.3.4 Skeletonization

Skeletonization is the process for reducing foreground regions in a binary image to a skeletal remnant that largely preserves the extent and connectivity of the original region while throwing away most of the original foreground pixels. After finding skeleton of hand end point and center point of skeleton is determined. Then line is drawn between end point and center point and reference line is drawn from the center point and angle between reference line and other line is calculated by using proper thresholding by counting the number of lines it is classified as number.

### 3.3.5 COM Determination

The center of area in binary images is the same as the center of the mass and it is computed as shown below:

$$\bar{x} = \frac{1}{A}\sum_{i=1}^{n}\sum_{j=1}^{m} jB[i, j] \quad \text{... (3.1)}$$

$$\bar{y} = \frac{1}{A}\sum_{i=1}^{n}\sum_{j=1}^{n} iB[i, j] \quad \text{... (3.2)}$$

Where: B is the matrix of size [n x m] representation of the region.
A is the area in pixels of the region

### 3.3.6 Circle construction

Circle is drawn whose radius is 0.65 of the farthest distance from the COG. Such a circle is likely to intersect all the fingers active in a particular gesture or "count."

By extracting a 1D binary signal by tracking the circle constructed in the previous step. Ideally the uninterrupted "white" portions of this signal correspond to the fingers or the wrist. By counting the number of zero-to-one (black-to-white) transitions in this 1D signal, and subtracting one (for the wrist) leads to the estimated number of fingers active in the gesture.

### 3.2.7 Fuzzy C-Means Clustering Algorithm

During the training operation, feature vectors ($x_1;x_2;x_3;$ : ; $x_n$) are calculated for all gesture images. Next step is the classification of these feature vectors based on similar data by using clustering technique. Generally K-means clustering (crisp) or Fuzzy-C-means clustering (fuzzy) techniques are preferred as the clustering techniques for classification of data vectors. But K-means clustering algorithm is sensitive to initial randomly selected cluster centers and moreover it does not find the optimal solution. On the contrary, Fuzzy-C-Means algorithm is not sensitive to initial cluster centers. During initialization, the assumptions made on membership functions are that the sum of membership values for each data set ($x_k$) in all cluster centers must be equal to 1 and these membership values must be positive[8], that is

$$\sum_{i=1}^{c} \mu_{ik} = 1, 1 \le \mathrm{k} \le \mathrm{n} \quad \dots (3.3)$$

The mathematical analysis of Fuzzy-C-Means algorithm is given below:

For a given data pattern set $X = fx1;x2;x3; : : : : ; xng$, which is obtained after the preprocessing of the gesture and C number of clusters, Fuzzy-C-Means algorithm classifies the data patterns by assigning them to cluster centers based on minimum Euclidian distance from each gesture feature vector to that cluster center. It is done by an iterative process in which for every iteration, cluster centers and membership vector matrices are updated until error objective function $J_m(U;V)$ reduces to zero. Here $U = \{\mu_{ik}\}$ is membership function space representing the degree of membership of *kth* data vector in the *ith* cluster and $V = (v_1; : : : : ; v_c)$ is the prototype weighted feature vector. "m" is the fuzzy weighting exponent , which plays a role in cluster validity of FCM algorithm and its value lies in the interval [1.5 2.5]. Generally the value of m will be chosen as 2. Steps in Fuzzy-C-Means clustering algorithm are:

1. For a given data set $X=(x_1;x_2, : : : ,x_n)$ and number of cluster centers "C", initialize membership function matrix $U(0)$, where $U = \{\mu_{ik}\}$.
2. Compute the cluster center vector $v_i^{(l)}$

3. Update $U^{(l)}=[\mu_{ik}^{(l)}]$ to $U^{(l+1)}=[\mu_{ik}^{(l+1)}]$

$$\mu_{ik}^{(l+1)} = \frac{1}{\sum_{j=1}^{C} \left( \frac{d_{ik}^{l}}{d_{jk}^{l}} \right)^{\frac{2}{m-1}}}, 1 \le \mathrm{i} \le \mathrm{C}, 1 \le \mathrm{k} \le \mathrm{n} \quad \dots (3.4)$$

If sum of squared error objective function ($J_m(U;V)$) converges then stop the iterations. Otherwise go to step (2).

$$J_m(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (\mu_{ik})^m \|x_k - v_i\|^2 \quad \dots (3.5)$$

# 4 RESULT

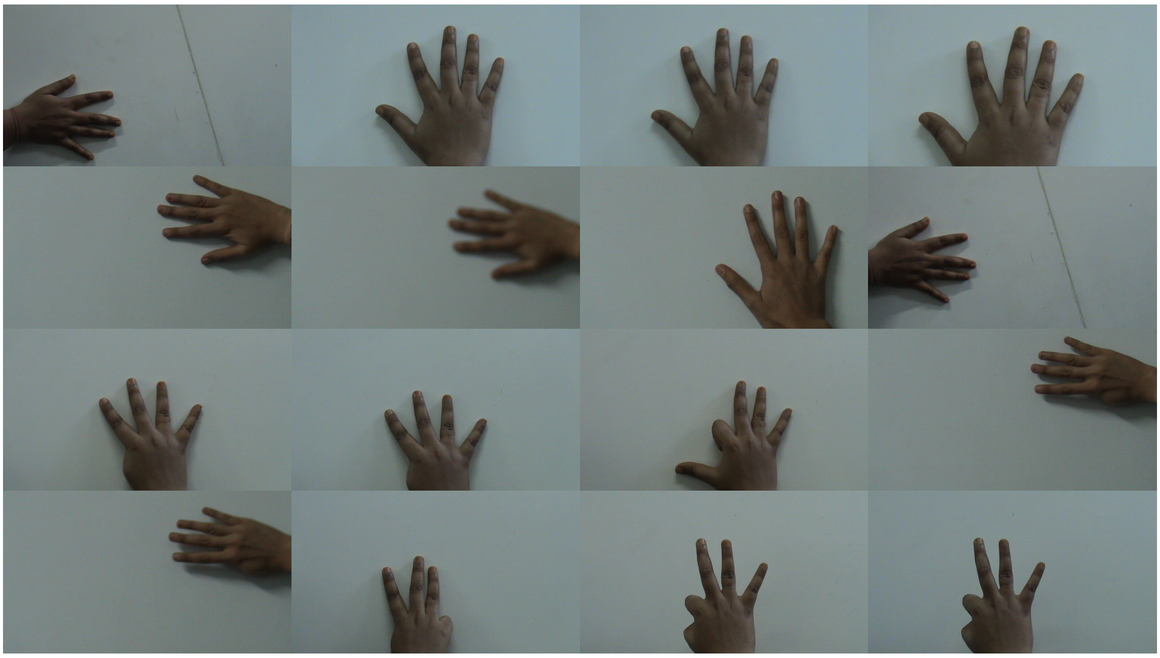
## 4.1 GENERAL

The system performs gesture recognition by extracting the feature from the hand in the input image.

## 4.2 PREPROCESSING

Preprocessing starts with obtaining the image, segmentation of the hand, morphological filtering and then feature extraction using boundary contour, skeletonization and circle construction.

After obtaining the images hand is separated from the rest of the images using color based segmentation. Skin color can be used as an important clue for isolating area of interest from the input image. The gesture used is shown in figure 4.1



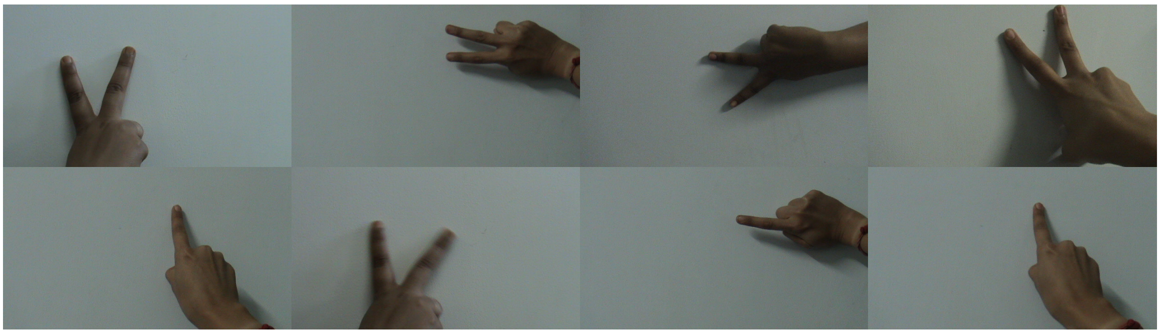


Figure 4.1 Gesture Database

The color histogram revealed that the distribution of skin-color of different people are clustered in the chromatic color space and a skin color distribution can be represented by a Gaussian model shown in figure 4.2

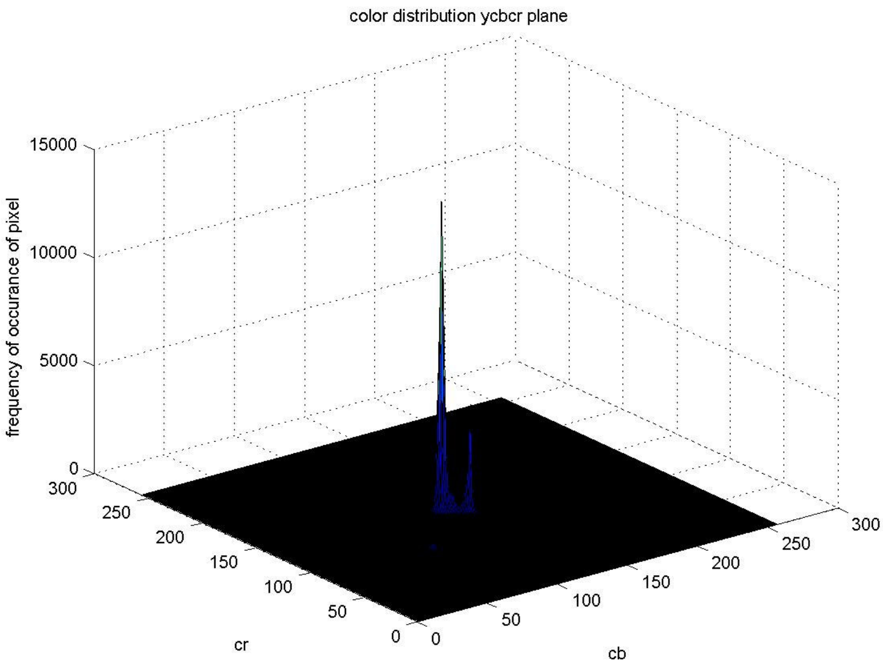



Figure 4.2 Color Distribution of skin

With this Gaussian fitted skin color model, we can now obtain the likelihood of skin for any pixel of an image shown in figure 4.3

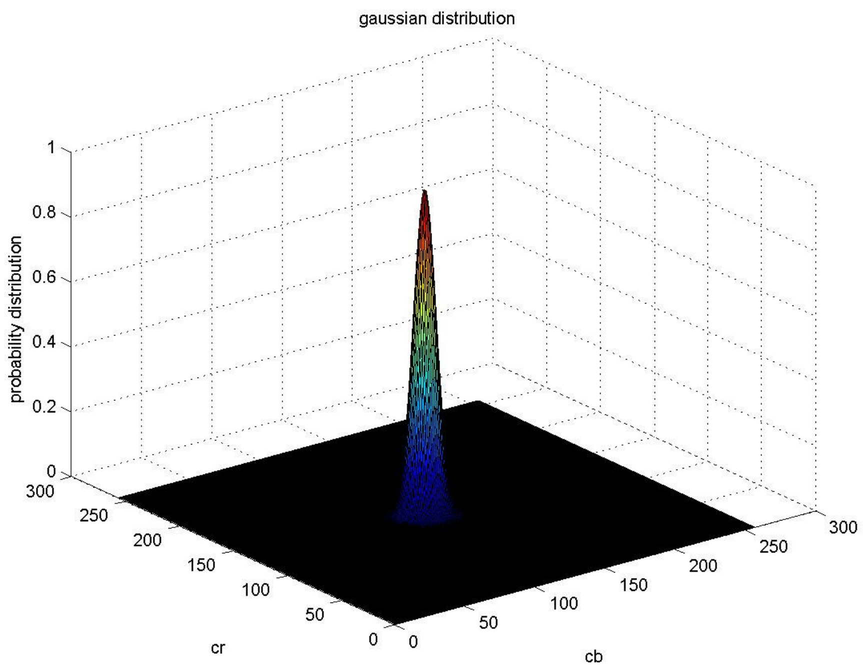
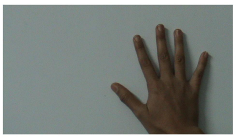


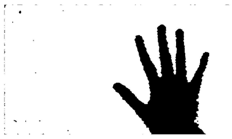

Figure 4.3 Gaussian distribution

The preprocessing of "Five" gesture is shown in figure 4.4



(a)

(b)



(c)

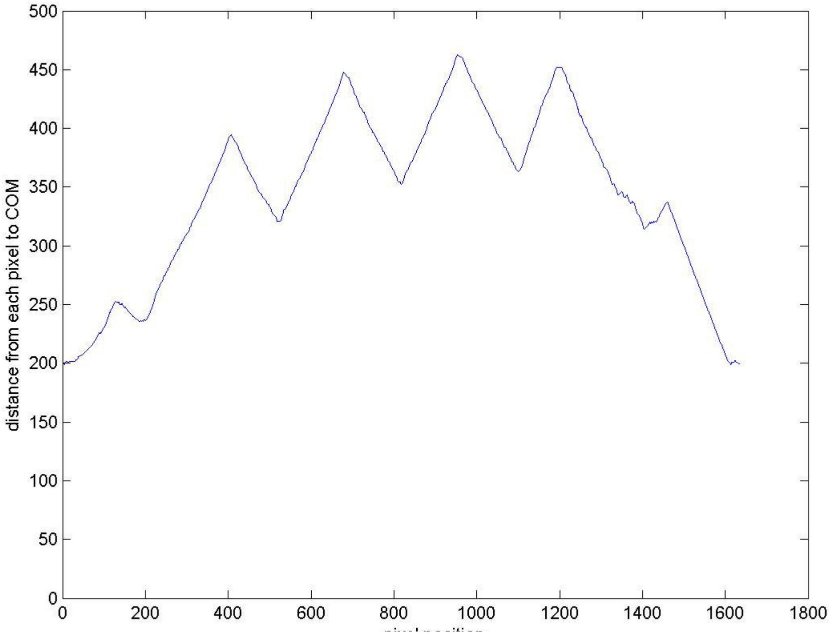
(d)

Figure 4.4 Preprocessing of "Five" Gesture. (a) Original Image. (b) Gray image. (c) Segmented image. (d) Binary image after inversion and morphological filtering

## 4.3 BOUNDARY CONTOUR

In this method for finding the boundary of hand first starting point is computed by bottom to top vertical scanning. Starting from that point, points along the edge of binary image is taken and saved in sequence, and at the same time computing the Euclidian distance between that point and the COM shown in figure 4.5.

(a)




(b)

Figure 4.5 (a) Boundary tracing of hand. (b) Distance from each pixel to COM.

The peaks (maxima) are the furthest point from the COM; they represent the positions of the tip of the fingers. From the graphs, first local maxima's are calculated by finding the first and second order difference and using the zero crossing property, then height of the maximum peak is determined. The minimum difference between the maximum and the closest minimum is computed. If that value is above 20% of the highest finger, it is encoded as '1' whilst peaks below that threshold are classified and encoded as '0'. By counting the total number of 1 gives the number represented by hand gesture.

## 4.4 SKELETON

The method used for skeltonization is thinning. To implement thinning, first, translate the origin of the structuring element (middle) to each possible pixel position in the image. If foreground and background pixels in the structuring element exactly match foreground and background pixels in the image, the image pixel underneath the origin of the structuring element is set to background. Otherwise it is left unchanged. In this project thinning method is used. MATLAB morphological function is used to find the skeletal representation of hand shown in figure 4.6. After finding skeleton pruning is applied in order to remove small branches from the skeleton. This is shown in figure 4.7

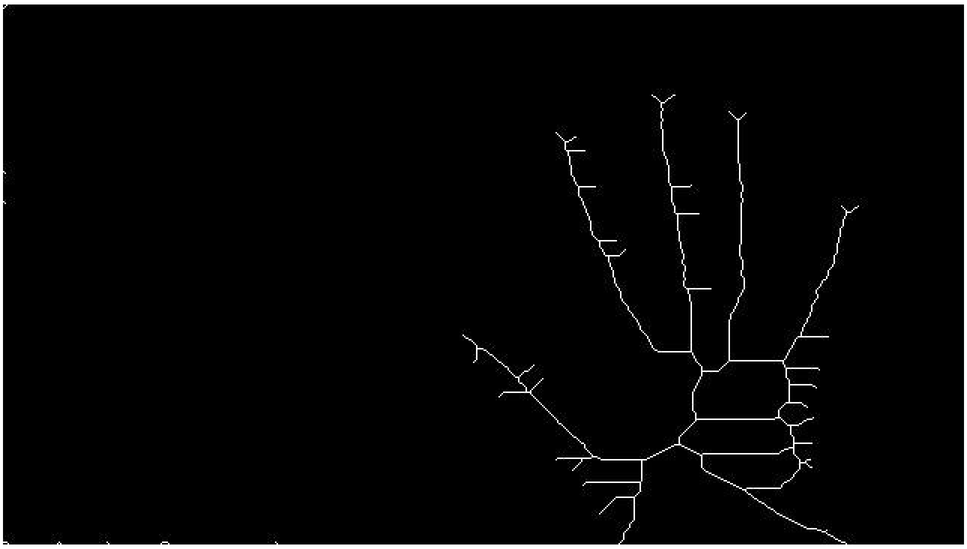


Figure 4.6 Skelton of hand

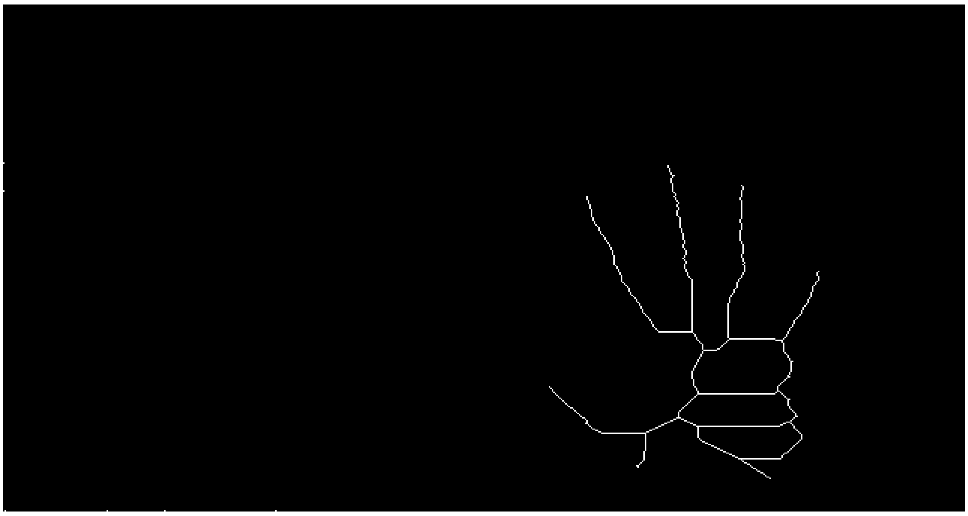


Figure 4.7 Skelton after pruning

After finding skeleton center point and end points are determined. After that line is drawn between center point and endpoint and reference line is drawn from center point. This is shown is figure 4.8

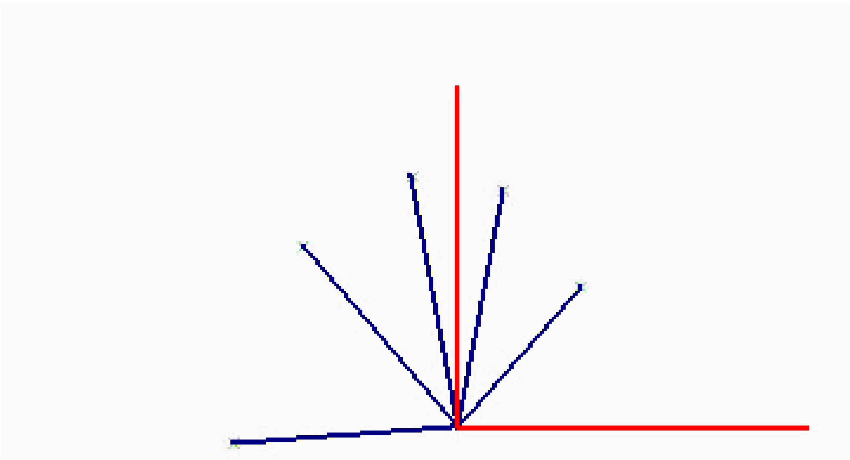


Figure 4.7 Lines between end points and center point

After drawing lines angle between different line and reference line is calculated if it lies in the range grater than -21 and less than 180 or less than -160 it is consider as one otherwise 0. the total number of one's thus obtained represents the numeral shown by the hand gesture.

## 4.5 CIRCLE CONSTRUCTION

Circle is drawn whose radius is 0.65 of the farthest distance from the COG. Such a circle is likely to intersect all the fingers active in a particular gesture or "count." Just to provide a visual flavor, Figure 4.8 demonstrates the execution of the steps described so far on a sample image.

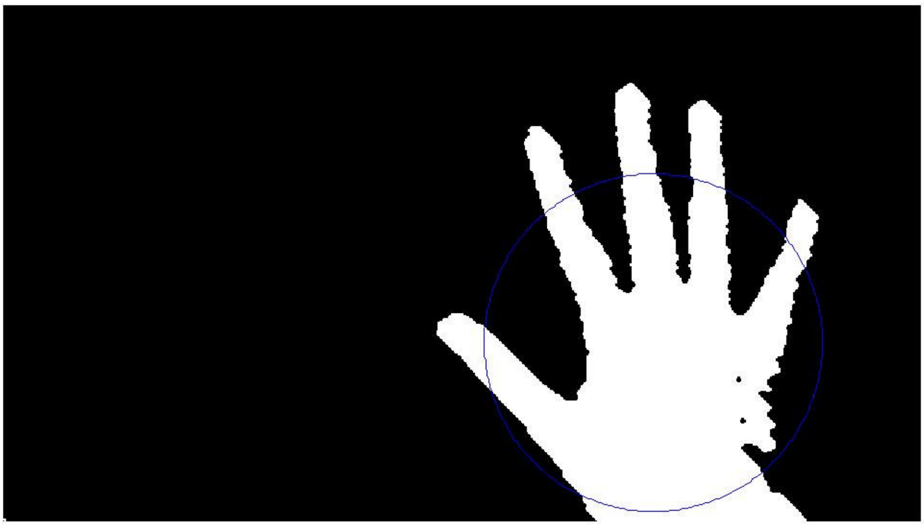


Figure 4.9 circle construction on hand binary image

Now extract a 1D binary signal by tracking the circle constructed in the previous step. Ideally the uninterrupted "white" portions of this signal correspond to the fingers or the wrist. By counting the number of zero-to-one (black-to-white) transitions in this 1D signal, and subtracting one (for the wrist) leads to the estimated number of fingers active in the gesture. Estimating the number of fingers leads to the recognition of the gesture combinations. This is shown in figure 4.9

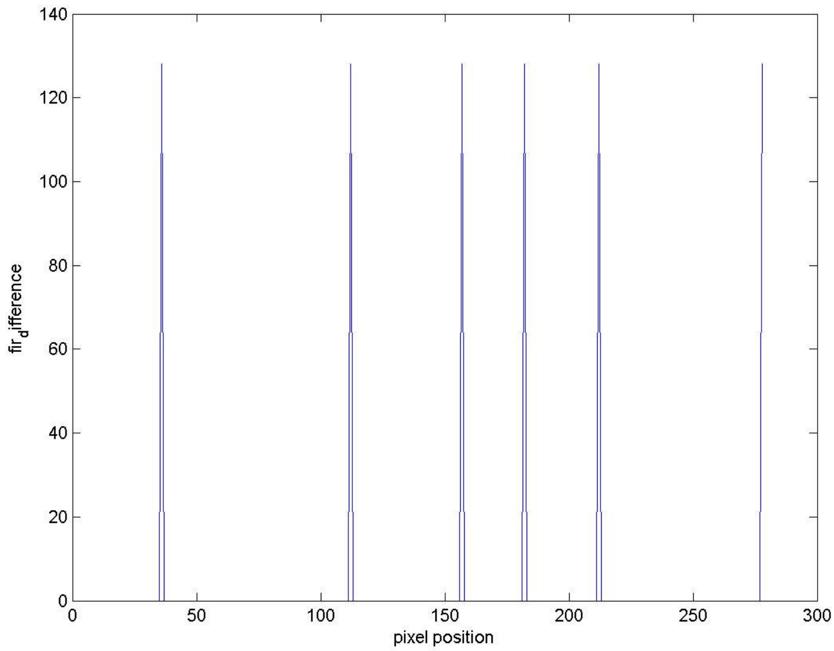



Figure 4.10 zero-one transition in the 1-D signal extracted

Table 4.1 Gesture recognition rate using boundary contour

| Gesture | Total Number of Gesture | Misclassified Gesture | Recognition Rate |
|---|---|---|---|
| One | 25 | 3 | 88% |
| Two | 25 | 4 | 84% |
| Three | 25 | 3 | 88% |
| Four | 25 | 4 | 84% |
| Five | 25 | 2 | 92% |

Table 4.2 Gesture recognition rate using Skeletonization

| Gesture | Total Number of Gesture | Misclassified Gesture | Recognition Rate |
|---|---|---|---|
| One | 25 | 4 | 84% |
| Two | 25 | 5 | 80% |
| Three | 25 | 6 | 76% |
| Four | 25 | 5 | 80% |
| Five | 25 | 4 | 84% |

Table 4.3 Gesture recognition rate using circle construction

| Gesture | Total Number of Gesture | Misclassified Gesture | Recognition Rate |
|---|---|---|---|
| One | 25 | 2 | 92% |
| Two | 25 | 3 | 88% |
| Three | 25 | 2 | 92% |
| Four | 25 | 2 | 92% |
| Five | 25 | 2 | 92% |

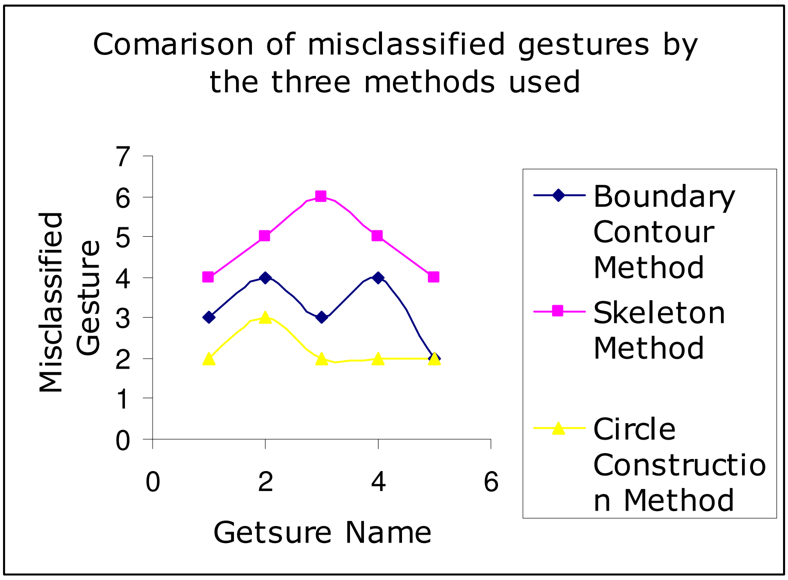



Figure 4.11 Misclassification rate of gesture (Total Gesture Taken is 25)

# 5 CONCLUSION and FUTURE WORK

## 5.1 CONCLUSION

The hand gesture recognition method for identification of numbers from the fingers of human hand was implemented by three different methods. The methods can be listed as Boundary Contour, Skeleton Method and Circle Construction Method. Each method was executed with a set of twenty five different images/gestures, for the recognition of the gestures mentioned above. Out of which, the classification rate varies between 76 - 92%.

The work related to classification of alphabets has been done up to feature extraction phase.

## 5.2 FUTURE WORK

Future work is to implement an algorithm for recognizing dynamic hand gestures and also to use it for recognizing new gesture. An obvious extension to the system, and one that is very easy to implement, is the simultaneous recognition of both hands in the input images.

## 5.3 DISCUSSION

Like Most of the other computer vision system, hand gesture recognition is highly context dependent. The system achieves its performance under a set of defined constraints. The segmentation module achieves tracking of skin region. Feature extraction step involves large set of constraints. Finally the classification is done on the basis of feature extracted.

## REFERENCES


1 Daniel Heckenberg, "MIME- A video based Hand Gesture Interface". The University of Queensland, Brisbane, Australia, October, 1999.

2 Kenny Teng, Jeremy Ng and Shirlene Lim, "Computer Vision Based Sign Language Recognition for Numbers". Carnegie Mellon University, spring 2004

3 E.di Benardo, L.Gonclaves, and p.perona, "Monocular Tracking of the Human Arm In 3d". Cipolla and Pentland, Ch. 9, pp.155-169

4 Robles Henry Chang and Ulises, "Face Detection". Standford University, May 25, 2000

5 Klimis Symeonidis , " Hand Gesture Recognition Using Neural Network". School of Electronic and Electrical Engineering, August 23, 2000.

6 Bogan Ionescu, Didir Coquin, Patrick lambert and Vasile Buzuloiu, "Dynamic Hand Gesture Recognition Using the Skeleton of the Hand". EURASIP journal on Applied Signal Processing 2005:13,2101-2109@ 2005 Hindawi publishing corporation

7 Asanterabi Malima, Erol Ozgur, and Mujdat Cetin " Fast Algorithm for Vision Based Hand Gesture Recognition for Robot Contro". Faculty of Engineering and Natural Sciences, Sabanc$_{1,}$ Tuzla, Istanbul, Turkey

8 V. S. Rao and C. Mahanta "Gesture based Robot Control". Department of Electronics and Communication Engineering, Indian Institute of Technology, Guwahati

9 Asutosh Saxena, Aditya Awathi, Vaibhav Vaish," SANKET- Interprets Your Hand Gestures". Indian Institute of Technology Kanpur, 2003.

10 Erdem Yoruk, Ender konuko glu, Bulent Sankur, Senior member, IEEE and Jerome Darbon." Shape-Based Hand Recognition IEEE Transaction on Image Processing

12 Nobuhiko Tanibata, Nobutaka Shimada and Yoshiaki Shirai, "Extraction of Hand Feature for Recognition of Sign Language". Computer-Controlled Mechanical System, Graduate School of Engineering, Osaka University, Australia

13 Yuan-Fei Cheng." Finding the Finger By the Boundary Vector Product". Department of Information Management, Meiho, Institute of Technology, Nei Pu, Ping Tung, Taiwan

14 Rafael C. Gonzalez, Richard E. Woods "Digital Image Processing". Second

Edition, Pearson Education

15 Hellman Asaf Hirshberg Doron “Gesture Recognition”.Israel institute of Technology, Department of Electrical Engineering The Vision and image sciences Laboratory

16 Leonid Sigal, Student Member, IEEE, Stan Sclaroff, Member, IEEE, and Vassilis Athitsos.’’ Skin Color-Based Video Segmentation under Time-Varying Illumination”, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 26, NO. 7, JULY 2004

17 Michael J. Jones and James M.Rehg. “Statastical Color Model with Application to Skin Detection”, Cambridge Research Laboratory, Compaq Computer corporation.

18 C. Maggioni and B. K¨ammerer, “Gesturecomputer - history, design and applications,” in Cipolla and Pentland [73], ch. 2, pp. 23–51

# List of Useful Sites


1 http://basic-eng.blogspot.com

2 "http://www.mathworks.com/matlabcentral/fileexchange/loadCategory.do?objectType=category&objectId=138&objectName=Morphology%20and%20Segmentation

3 "http://www.mathworks.com/access/helpdesk/help/toolbox/images/index.html?/access/helpdesk/help/toolbox/images/f8-22867.html&http://www.google.co.in/search?q=color+image+processing&hl=en&lr=&start=30&sa=N

4 "http://www-cs-students.stanford.edu/~robles/ee368/main.html"