# Intelligent Sign Language Recognition Using Image Processing

## Sawant Pramada[1], Deshpande Saylee [2,] Nale Pranita[3,] Nerkar Samiksha[4] Mrs.Archana S. Vaidya[5]

[1,2,3,4,5]GES's R. H. Sapat College of Engineering, Management Studies and Research, Nashik
(M.S.), INDIA
[5]Asst. Prof. Department of Computer Engineering

***Abstract*** *: Computer recognition of sign language is an important research problem for enabling communication with hearing impaired people. This project introduces an efficient and fast algorithm for identification of the number of fingers opened in a gesture representing an alphabet of the Binary Sign Language. The system does not require the hand to be perfectly aligned to the camera. The project uses image processing system to identify, especially English alphabetic sign language used by the deaf people to communicate. The basic objective of this project is to develop a computer based intelligent system that will enable dumb people significantly to communicate with all other people using their natural hand gestures. The idea consisted of designing and building up an intelligent system using image processing, machine learning and artificial intelligence concepts to take visual inputs of sign language's hand gestures and generate easily recognizable form of outputs. Hence the objective of this project is to develop an intelligent system which can act as a translator between the sign language and the spoken language dynamically and can make the communication between people with hearing impairment and normal people both effective and efficient. The system is we are implementing for Binary sign language but it can detect any sign language with prior image processing*

***Keywords:*** *Artificial Intelligence, Binary sign Language, Image Processing, Machine learning, Template Matching.*

## I.      INTRODUCTION

Dumb people are usually deprived of normal communication with other people in the society. It has been observed that they find it really difficult at times to interact with normal people with their gestures, as only a very few of those are recognized by most people. Since people with hearing impairment or deaf people cannot talk like normal people so they have to depend on some sort of visual communication in most of the time.

Sign Language is the primary means of communication in the deaf and dumb community. As like any other language it has also got grammar and vocabulary but uses visual modality for exchanging information. The problem arises when dumb or deaf people try to express themselves to other people with the help of these sign language grammars. This is because normal people are usually unaware of these grammars. As a result it has been seen that communication of a dumb person are only limited within his/her family or the deaf community.

The importance of sign language is emphasized by the growing public approval and funds for international project. At this age of Technology the demand for a computer based system is highly demanding for the dumb community. However, researchers have been attacking the problem for quite some time now and the results are showing some promise. Interesting technologies are being developed for speech recognition but no real commercial product for sign recognition is actually there in the current market.

The idea is to make computers to understand human language and develop a user friendly human computer interfaces (HCI). Making a computer understand speech, facial expressions and human gestures are some steps towards it. Gestures are the non-verbally exchanged information. A person can perform innumerable gestures at a time. Since human gestures are perceived through vision, it is a subject of great interest for computer vision researchers. The project aims to determine human gestures by creating an HCI. Coding of these gestures into machine language demands a complex programming algorithm. In our project we are focusing on Image Processing and Template matching for better output generation.

## II.      LITERATURE SURVEY

Not many Researches have been carried out in this particular field, especially in Binary Sign Language Recognition. Few researches have been done on this issue though and some of them are still operational, but nobody was able to provide a full fledged solution to the problem. Christopher Lee and Yangsheng Xu developed a glove-based gesture recognition system that was able to recognize 14 of the letters from the hand alphabet, learn new gestures and able to update the model of each gesture in the system in online mode, with a

rate of 10Hz. Over the years advanced glove devices have been designed such as the Sayre Glove, Dexterous Hand Master and Power Glove [1].

The most successful commercially available glove is by far the VPL Data Glove [2]. It was developed by Zimmerman during the 1970's. It is based upon patented optical fiber sensors along the back of the fingers. Star-ner and Pentland developed a glove-environment system capable of recognizing 40 signs from the American Sign Language (ASL) with a rate of 5Hz.

Another research is by Hyeon-Kyu Lee and Jin H. Kim presented work on real-time hand-gesture recognition using HMM (Hidden Markov Model). Kjeldsen and Kendersi devised a technique for doing skin-tone segmentation in HSV space, based on the premise that skin tone in images occupies a connected volume in HSV space. They further developed a system which used a back-propagation neural network to recognize gestures from the segmented hand images[1].

Etsuko Ueda and Yoshio Matsumoto presented a novel technique a hand-pose estimation that can be used for vision-based human interfaces, in this method, the hand regions are extracted from multiple images obtained by a multi viewpoint camera system, and constructing the "voxel Model"[6] . Hand pose is estimated. Chan Wah Ng, Surendra Ranganath presented a hand gesture recognition system, they used image furrier descriptor as their prime feature and classified with the help of RBF network. Their system's overall performance was 90.9%. Claudia Nolker and Helge Ritter presented a hand gesture recognition modal based on recognition of finger tips, in their approach they find full identification of all finger joint angles and based on that a 3D modal of hand is prepared and using neural network.

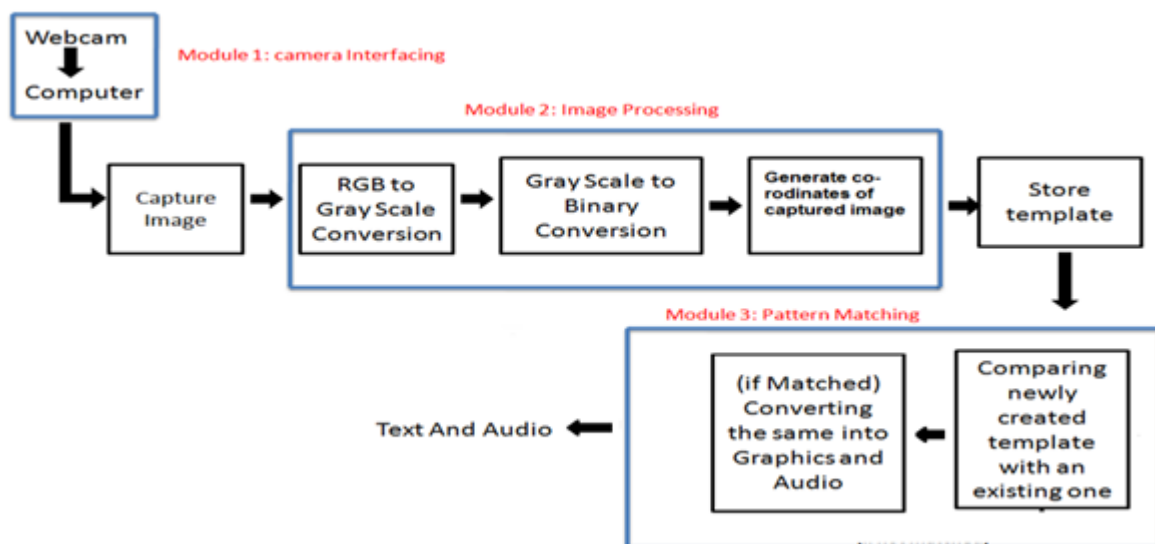## III.       SYSTEM ARCHITECTURE



Fig 1. System diagram for proposed system.

Fig 1 shows the overall idea of proposed system. The system consists of 4 modules. Image is captured through the webcam. The camera is mounted on top of system facing towards the wall with neutral background. Firstly, the captured Colored image is converted into the gray scale image which intern converted into the binary form. Coordinates of captured image is calculated with respect to X and Y coordinates. The calculated coordinates are then stored into the database in the form of template. The templates of newly created coordinates are compared with the existing one. If comparison leads to success then the same will be converted into audio and textual form. The system works in two different mode i.e. training mode and operational mode.
Training mode is part of machine learning where we are training our system to accomplish the task for which it is implemented i.e. Alphabet Recognition.

### 3.1 Camera Interfacing and Image Acquisition
The Camera Interface block is the hardware block that interfaces and provides a standard output that can be used for subsequent image processing.

### 3.1.1 Camera Orientation:
It is important to carefully choose the direction in which the camera points to permit an easy choice of background. The two realistic options are to point the camera towards a wall or towards the floor (or desktop).

However since the lighting was single overhead bulb, light intensity would be higher and shadowing effects least if the camera was pointed downwards.

### 3.1.2 Camera Specifications:

We are using Intex Night Vision 16 MP Webcam. The Intex 16 MP Webcam comes with Night Vision feature. This webcam gives clear video imaging and can work even in the darkness. The night vision gives good images in the dark also. The upper portion of the webcam is Movable depend on the need. The resolution of captured image is 640x480 having frame rate up to 30fps and The Format of image is RGB 24, 1420.

### 3.2 Image Processing

### 3.2.1 RGB Color Recognition

Basically, any color image is a combination of red, green, blue colors. An important trade-off when implementing a computer vision system is to select whether to differentiate objects using colour or black and white and, if colour, to decide what colour space to use (red, green, blue or hue, saturation, luminosity)[1]. For the purposes of this project, the detection of skin and marker pixels is required, so the colour space chosen should best facilitate this. The camera available permitted the detection of colour information. Although using intensity alone (black and white) reduces the amount of data to analyze and therefore decreases processor load it also makes differentiating skin and markers from the background much harder (since black and white data exhibits less variation than colour data). Therefore it was decided to use colour differentiation. Further maximum and minimum HSL pixel colour values of a small test area of skin were manually calculated. These HSL ranges were then used to detect skin pixels in a subsequent frame (detection was indicated by a change of pixel colour to white). But Hue, when compared with saturation and luminosity, is surprisingly bad at skin differentiation (with the chosen background) and thus HSL shows no significant advantage over RGB. Moreover, since conversion of the colour data from RGB to HSL took considerable processor time it was decided to use RGB.[3]

We will take the color image. Then make required portion of image as white by using Thresholding technique( as explained below) and garbage part that is background as black. Then we get black and white image and it is compared with the stored template.

### 3.2.1.1 Color image to Binary image conversion

To convert any color to a grayscale representation of its luminance, first one must obtain the values of its red, green, and blue (RGB) primaries. Grayscale or grayscale digital image is an image in which the value of each pixel is a single sample, that is, it carries only intensity information. Images of this sort, also known as black and white, are composed exclusively of shades of gray, varying from black at the weakest intensity to white at the strongest. A binary image is a digital image that has only two possible values for each pixel. Typically the two colors used for a binary image are black and white though any two colors can be used. The color used for the object in the image is the foreground color while the rest of the image is the background colour. Until now a simple RGB bounding box has been used in the classification of the skin and marker pixels.[4]

### 3.2.2 Thresholding

Thresholding is the simple method of image segmentation[1]. In this method we convert the RGB image to Binary image. Figure 2-5 shows the details of  image processing . Binary image is digital image and has only two values (0 or 1). For each pixel typically two colors are used black and white though any two colors can be used. Here, the background pixels are converted into black color pixels and pixels containing our area of interest are converted into white color pixels. It is nothing but the preprocessing.
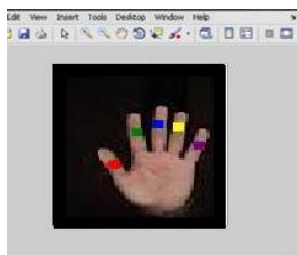


Fig.2  Original Image            Fig.3  Binary Image
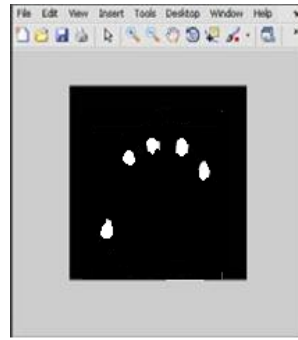
Fig. 4 Binary Image(Mask image After Thresholding)   Fig.5 only the area of interest is preserved
is discarded.(using color filters)

### 3.2.3Coordinate Mapping

In the previous step only the area containing marker color bands are preserved for further processing and rest of the portion of image are converted into black color pixels this is shown in figure 6. This task of converting color band pixels into the white color pixels is accomplished by setting the values of RGB color in filter. After getting the marker pixels that are now highlighted as a white color pixels, coordinates of that area for each color is generated. The newly generated coordinates are the compared with the stored coordinates in the database for the purpose of output generation using pattern matching technique explained in the next section.
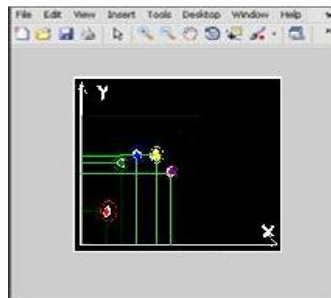


Fig. 6 Coordinate Mapping

It is likely that the detection system will be subjected to varying lighting conditions (for example, due to time of day or position of camera relative to light sources). Therefore it is likely that an occasional recalibration will have to be performed. The calibration technique is discussed below:

### 3.2.3.1 Color Calibration

In order to automatically calculate the colour ranges (1), an area of the screen was demarcated for calibration (2). It was then a simple matter to position the hand or marker (color rings) within this area and then scan it to find the maximum and minimum RGB values of the ranges(3).

A formal description of the initial calibration method is as follows: The image is a 2D array of pixels:

$$\vec{I}(x,y) = \begin{pmatrix} r(x,y) \\ g(x,y) \\ b(x,y) \end{pmatrix}$$

. . . . . . . . . . . . . (1)

The calibration area is set of 2D points:

$$\vec{J} = \{\vec{x}_1 \ldots \vec{x}_n\}$$
$$\text{where } \vec{x}_i = (x,y)$$

. . . . . . . . . (2)

The color ranges can then be defined for this area:

$$r_{max} = \max_{\bar{x} \in J} r(\bar{x}) \quad r_{min} = \min_{\bar{x} \in J} r(\bar{x})$$
$$g_{max} = \max_{\bar{x} \in J} g(\bar{x}) \quad g_{min} = \min_{\bar{x} \in J} g(\bar{x})$$
$$b_{max} = \max_{\bar{x} \in J} b(\bar{x}) \quad b_{min} = \min_{\bar{x} \in J} b(\bar{x}) \quad \ldots\ldots\ldots (3)$$

A formal description of skin detection is then as follows,

The skin pixels are those pixels(r, g, b) such that,

$$((r \geq r_{min}) \,\&\, (r \leq r_{max})) \,\&$$
$$((g \geq g_{min}) \,\&\, (g \leq g_{max})) \,\&$$
$$((b \geq b_{min}) \,\&\, (b \leq b_{max}))$$

Call this Predicate S(r, g, b)

The set of all skin pixels location is then:

$$L = \left\{ \bar{x} \mid S(r(\bar{x}), g(\bar{x}), b(\bar{x})) = 1 \right\}$$

Using this method skin pixels were detected at the rate of 15fps on 2.00 GHz laptop

**3.3 Pattern Matching Algorithm**
　　　In this method, the input image after processing is set to the pixel values (3) of each color to be used such as Red_new (Rx, Ry), Green_new (Gx, Gy), Blue_new (Bx, By), Purple_new (Px, Py), Yellow_new (Yx, Yy).Pixel values comprise of the minimum and maximum values of each color pixel or can be called as co-ordinates. The generated values of these co-ordinates will be then compared with the values stored in the templates stored in the database. To obtain these values the general idea is firstly to find the area of each color pixel and the coordinates (Yx,Yy) by using the equation:
　　　Area= count number of white pixels obtained by Thresholding.

$$X = \frac{\text{Moment in X direction}}{\text{Area}}$$

$$Y = \frac{\text{Moment in Y direction}}{\text{Area}}$$

　　　Each newly generated pixel value then gets compared with the previously stored template value in the database. Algorithm proceeds until the comparison leads to success or failure. If algorithm returns positive result then the sign will be converted into corresponding text and audio if comparison results into failure then the proper error message will be displayed on the screen.

**3.4 Text to speech conversion**
　　　This part comes under Artificial Intelligence[4] where once the template matching operation becomes successful the matched image is then translated into text and audio format. For this purpose, predefined methods are used for conversion.

## IV.　　ALPHABET RECOGNITION
Following table 1 shows the values assigned to each finger.[5] Binary Alphabet calculation: It is possible to display total $(2^5 - 1)$ i.e.31 gestures using the fingers of a single hand, and $(2^{10} - 1)$ i.e. 1023 gestures if both hands are used.

Table.1 Values assigned to each finger

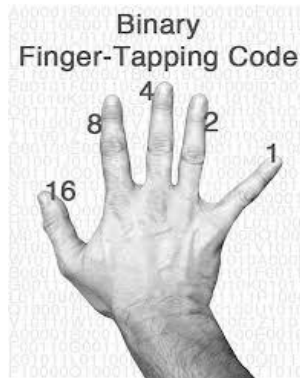| | Pinky | Ring | Middle | Index | Thumb |
|---|---|---|---|---|---|
| Power of two | $2^4$ | $2^3$ | $2^2$ | $2^1$ | $2^0$ |
| Value | 16 | 8 | 4 | 2 | 1 |

Fig. 7  Binary finger tapping tool

Figure 7 show binary finger tapping tool showing the significant values assigned to fingers by referring to gesture table 2 shows code of each alphabet.

Table 2. Alphabets code

| SR NO | RED | GREEN | BLUE | YELLOW | PURPLE | ALPHABETS |
|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 1 | A |
| 2 | 0 | 0 | 0 | 1 | 0 | B |
| 3 | 0 | 0 | 0 | 1 | 1 | C |
| 4 | 0 | 0 | 1 | 0 | 0 | D |
| 5 | 0 | 0 | 1 | 0 | 1 | E |
| 6 | 0 | 0 | 1 | 1 | 0 | F |
| 7 | 0 | 0 | 1 | 1 | 1 | G |
| 8 | 0 | 1 | 0 | 0 | 0 | H |
| 9 | 0 | 1 | 0 | 0 | 1 | I |
| 10 | 0 | 1 | 0 | 1 | 0 | J |
| 11 | 0 | 1 | 0 | 1 | 1 | K |
| 12 | 0 | 1 | 1 | 0 | 0 | L |
| 13 | 0 | 1 | 1 | 0 | 1 | M |
| 14 | 0 | 1 | 1 | 1 | 0 | N |
| 15 | 0 | 1 | 1 | 1 | 1 | O |
| 16 | 1 | 0 | 0 | 0 | 0 | P |
| 17 | 1 | 0 | 0 | 0 | 1 | Q |
| 18 | 1 | 0 | 0 | 1 | 0 | R |
| 19 | 1 | 0 | 0 | 1 | 1 | S |
| 20 | 1 | 0 | 1 | 0 | 0 | T |
| 21 | 1 | 0 | 1 | 0 | 1 | U |
| 22 | 1 | 0 | 1 | 1 | 0 | V |
| 23 | 1 | 0 | 1 | 1 | 1 | W |
| 24 | 1 | 1 | 0 | 0 | 0 | X |
| 25 | 1 | 1 | 0 | 0 | 1 | Y |
| 26 | 1 | 1 | 0 | 1 | 0 | Z |



| Fig 8.(a) | Fig 8.(b) | Fig 8.(c) |

Fig.8  Gesture of alphabet G from table 2.

Fig 8.a shows the original image where values of red and green color are 0 and rest of the color values are 1.
Fig 8.b shows the binary image after Thresholding and Fig 8.c shows image in which only the marker bands are highlighted.

# V.      CONCLUSION

Our project aims to bridge the gap by introducing an inexpensive computer in the communication path so that the sign language can be automatically captured, recognized and translated to speech for the benefit of blind people. In the other direction, speech must be analyzed and converted to either sign or textual display on the screen for the benefit of the hearing impaired.

## 4.1 Future scope

Motivations to carry out further research in order to develop enhanced version of the proposed system. System would be able to communicate in both directions i.e. It will have the capability to translate normal languages to hand gestures successfully. The image processing part of the system will also be modified to work with every possible environment.. A challenge will be to recognize signs that involve motion.

# REFERENCES

[1]     Christopher Lee and Yangsheng Xu, "Online, interactive learning of gestures for human robot interfaces" Carnegie Mellon University, the Robotics Institute, Pittsburgh, Pennsylvania, USA, 1996

[2]     Richard Watson, "Gesture recognition techniques", Technical report, Trinity College, Department of Computer Science, Dublin, July, Technical Report No. TCD-CS-93-11, 1993

[3]     Ray Lockton, "Hand Gesture Recognition using computer Vision",4th year project report, Ballilol College , Department of Engineering Science, Oxford University,2000

[4]     Ms. Rashmi D. Kyatanavar, Prof. P. R. Futane, "Comparative Study of Sign Language Recognition Systems", Department of Computer Engineering, Sinhgad College of Engineering,  Pune, India International Journal of Scientific and Research Publications, Volume 2, Issue 6, June 2012 ISSN 2250-3153

[5]     International Multi Conference of Engineers and Computer Scientists 2009",  Hong Kong , Vol I IMECS 2009, March 18 - 20, 2009

[6]     Etsuko Ueda, Yoshio Matsumoto, Masakazu Imai, Tsukasa Ogasawara, "Hand Pose Estimation for Vision Based Human Interface", IEEE Transactions on Industrial Electronics,Vol.50,No.4,pp.676-684,2003.

**Books:**

[7]      "digital image processing" (2nd Edition) Rafael C. Gonzalez (Author), Richard E. Woods (Author) Publication Date: January 15, 2002 | ISBN-10: 0201180758 | ISBN-13: 978-0201180756