ISSN No:-2456-2165

Hand Gesture to Text and Speech Conversion

Ohnmar Win
Department of Electronic Engineering
Mandalay Technological University,
Mandalay, Myanmar

Abstract:- A dumb person always uses gestures to convey his ideas to others. But it is always difficult for a normal person to understand this sign language. This article presents a solution that automatically recognizes hand gestures and converts them into text and language so that people with intellectual disabilities can easily communicate with normal people. The images of the hand gestures are acquired using the built-in LAPTOP camera or the webcam through a color detection technique. The gestures of the hand are detected to realize the different gesture of the hand. The system uses image processing technology for capturing and converting gestures. The hand gesture detections are performed in the operating system and provide speech output via a speaker. Based on recognized hand gestures, the recorded sound will be played. In the proposed work, the captured images are processed via MATLAB on PC and converted to screen text and speech by loudspeaker.

Keywords:- *Hand Gesture Recognition, Speech, Text, MATLAB.*

I. INTRODUCTION

One of the major problems facing our society is that people with disabilities are finding it hard to communicate with normal people without interpreter. In the recent years, there has been a rapid increase in the number of hearing impaired and speech disabled victims due to birth defects, oral diseases and accidents. The normal person seldom understands that a deaf-dumb person want to speak. So the deaf-dumb person shows gestures for his/her needs. Dumb persons have their own language to communicate with us. Generally they use sign language to communicate with other people. But they find difficulty in communicating with others who don't understand sign language [1].

The sign language translation system translates the normal sign language for example the American Sign Language, the British Sign Language, the Japanese Sign Language to speech and hence makes the communication between normal person and dumb people easier. When deaf-dumb person want to communicate with other people it is needed an experienced and qualified interpreter to understand their sign language. But it is a very difficult task and also unaffordable to get an interpreter in every time. So the implementation of the system that automatically recognizes the sign language is necessary to provide a better platform for the interaction of the deaf, dumb and blind people with the rest of the world without an interpreter.

The approaches present can be mainly divided into "Data-Glove based" and "Vision Based" approaches. The Data-Glove based methods use sensor devices for digitizing hand and finger motions into multi-parametric data. The extra sensors make it easy to collect hand configuration and movement. However, the devices are quite expensive and bring much cumbersome experience to the users. In contrast, the Vision Based methods require only a camera, thus realizing a natural interaction between humans and computers without the use of any extra devices [1]. In this paper, we introduce "Vision Based Hand Gesture to Speech and Text Conversion System".

II. PROPOSED SYSTEM

The deaf-dumb person use sign language for communicating and understating by other people. In sign language, It is needed the hand gestures, facial expression and body movements are needed for understanding one word. Many research works related to Sign languages have been done for one character at once. So, facial expression detection, body movement detection and hand gesture detection are necessary to provide a better system that automatically recognizes the sign language.

In the proposed system, one gesture can be used for one word. The words are predefined for each gesture. At the instant we considered the ten words. So the user need to know which gesture is used for which word that defined for this proposed system. The system is designed with a minimized cost and higher accuracy. The proposed system block diagram Fig. 1 consists of mainly three processes i.e. Image acquisition, gesture recognition and conversion to text and speech.

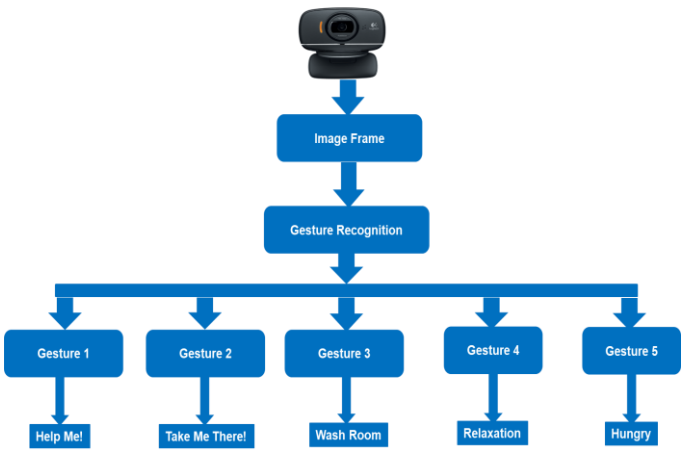


Fig 1:- Overall System Block Diagram

There are different steps used for the hand gesture recognition and text and speech conversion system. The following steps were implemented.

- *Image Acquisition:* Image acquisition is the initial stage of the vision system which is obtained by taking the snap of the image in the inbuilt camera or an external camera.
- *RGB to GRAY Conversion:* Capture image is converted to RGB to gray
- *Detection of Hand Region:* The concept of R, G, B color channel extraction and thresholding (limit) is used for image segmentation.
- *COG:* The center of gravity is calculated. Then the detected fingertips are named by numbered.
- *Finger Tip Detection:* Finding Edge and Convex Hull Vertices
- *Defined Conditions for Hand Gesture:* **S**et the conditions to corresponding words as per hand gestures.

A. Image Acquisition

Hand gesture image is captured from the built-in LAPTOP camera in real time. So the user needs to make gestures by placing the hand in front of the webcam. Images are continuously captured and converted into frames in this system. The user must use the red glove for conveniently the using of application.

B. RGB to Gray Conversion

The captured image is RGB format: Red, Green and Blue color. Each color plane is extracted for segmenting the hand region more exactly. And then it is converted into grayscale image.

C. Detection of Hand Region

Color detection algorithm was used that detects the color region from the input image as the background may consist of many other things along with the hand region. The color suppressed channel is subtracted from the Gray-scale Image for detecting the red color object. This created hand gesture image as a patch of gray surrounded by black space. Finally that image is converted into a binary image by using thresholding method.

D. Center of Gravity

Center of gravity (COG) is calculated from big contour of the binary image. The center of gravity is simply described by the spatial moments of order m_{10} and m_{01} divided by the zero order member m_{00} of the struct CV Moments.

First order moment,

$$m_{1,0} = \iint dxdyf(x,y), \quad (1)$$

$$m_{0,1} = \iint dxdyf(x,y), \quad (2)$$

$$x_c = \frac{m_{10}}{m_{00}} \quad (3)$$

$$y_c = \frac{m_{00}}{m_{10}} \quad (4)$$

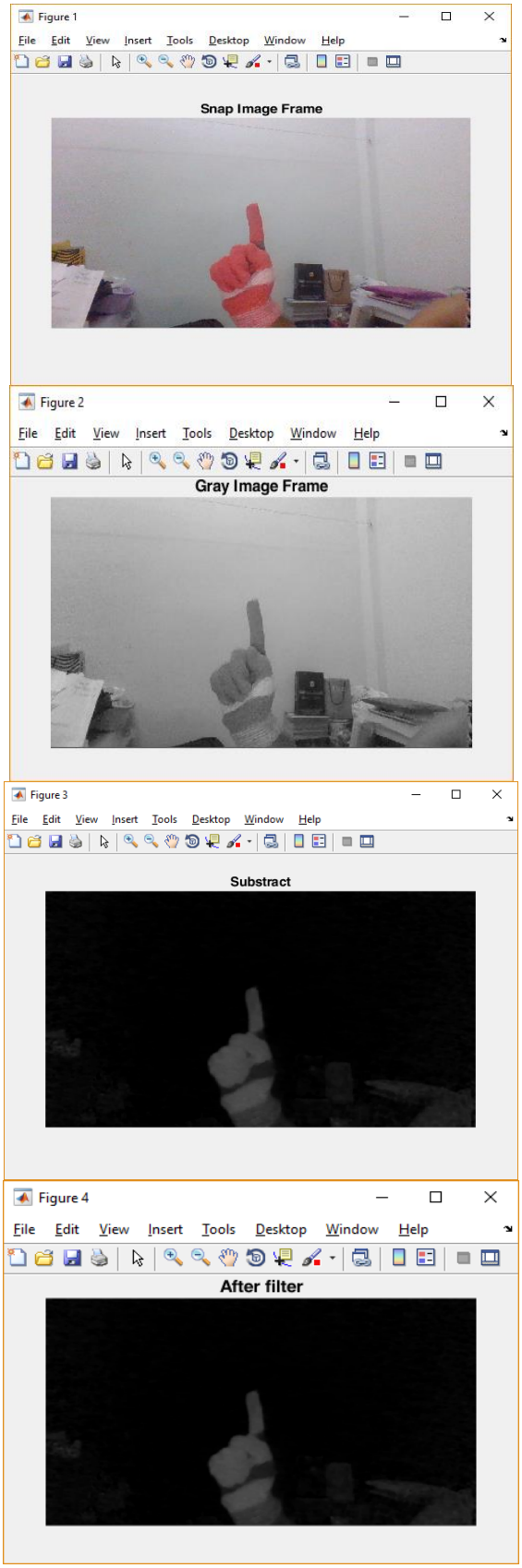
E. Convex Hull Algorithm

The algorithm takes in a binary image produced in which de-noising procedures have been applied. To find the convex hull, the global maximum point is used as the first convex hull vertex. Bresenham lines are then drawn from the first vertex to the right edge of the bounding box with the purpose of looking for intersection point with the hand .By using Jarvis March algorithm, the intersection point is the second convex hull vertex found and the process is repeated until all four bounding box edges have been processed .In this step, the white pixels of the image are scanned once to find the four global extreme values, M_x, m_x, M_y, m_y and a global maximum point.

$$H(x,y) = \int_{0,background}^{255,hand\ pixels} m_x \leq x \leq M_x, m_y \leq y \leq M_y$$

III. RESULTS

First the system will take the input from the user by using webcam i.e., the original image. Images are continuously captured as video file and then image frame is extracted from video. In the next step, the capture image frame will be converted into gray scale in parallel approach. Next important step will be color detection for hand region extraction. Image segmentation for hand extraction was the color based detection that the user has been worn the red groove for this system. An image contains the detected hand region is obtained. After recognizing hand it is filtered by using median filter for smoothing purpose and converted into a binary image by using thresholding method. Finally the segmented hand gesture image is with the skin regions represented in white color and all other non-skin regions in black. This gives the segmented hand region and this is the region of interest. In next step, the box surrounding the contour is used to receive a center.



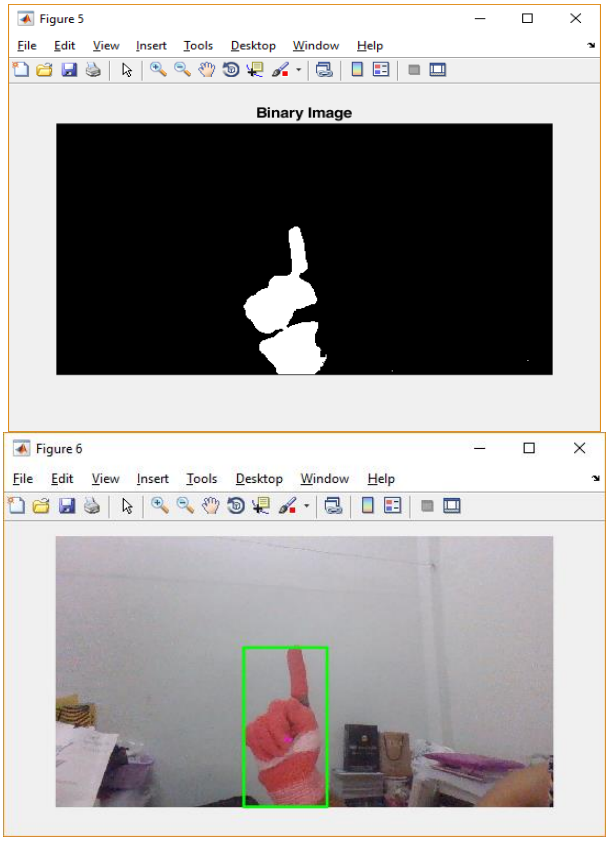


Fig 2:- Step by Step Testing Results

In this system spatial moments is used to obtain the Centre of Gravity of an input binary image. Center of gravity (COG) is calculated from big contour of the binary image. The next step is Finger Tips Detection for counting fingers and knowing hand gesture. Each hand gesture was mapped to a particular action as mentioned above, then the corresponding voice will be generated. The text are viewed in the display is shown in the Fig 3 to 6. The voice will be heard through the speaker.

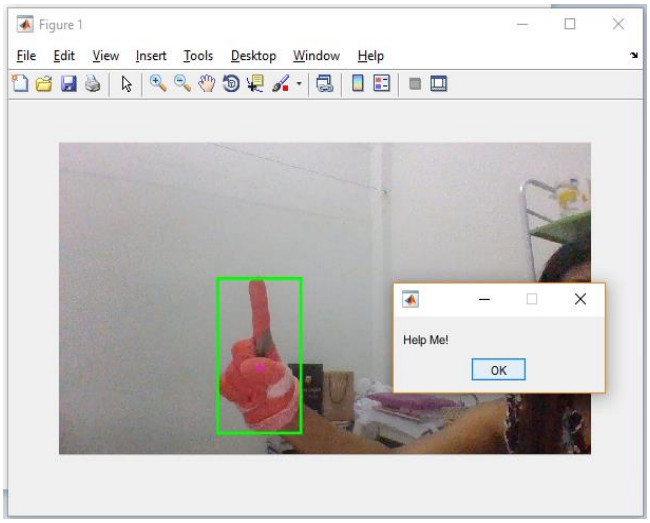


Fig 3:- Testing Result for Gesture 1

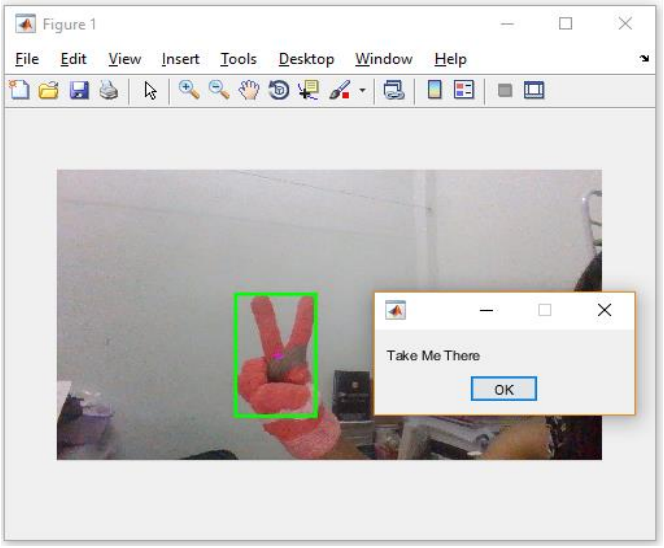


Fig 4:- Testing Result for Gesture 2

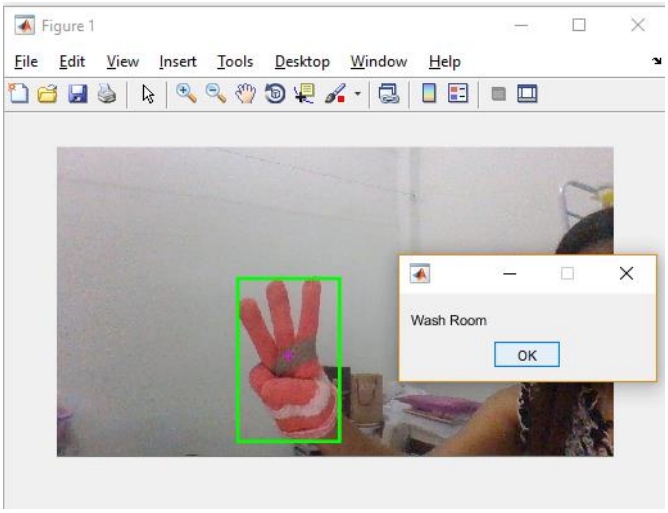


Fig 5:- Testing Result for Gesture 3

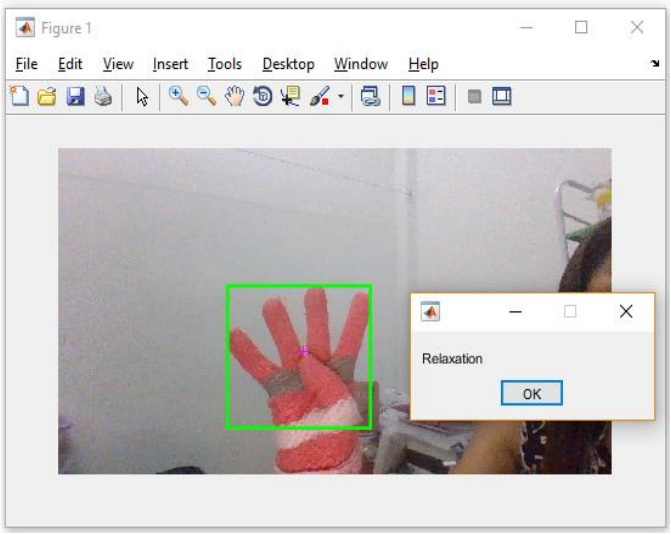


Fig 6:- Testing Result for Gesture 4

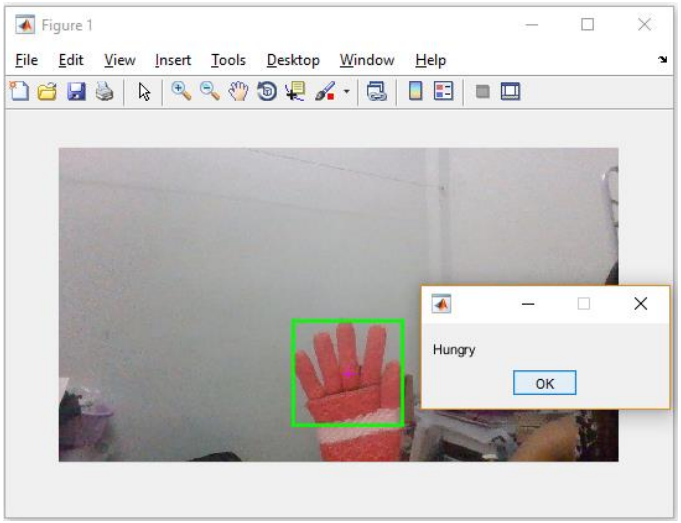


Fig 7:- Testing Result for Gesture 5

IV. CONCLUSION

In this work, a vision based hand gesture recognition system using MATLAB software for automatic conversion of text and speech has been presented. On the whole, the solution aims to provide aid to those in need thus ensuring social relevance. By using this system the deaf-dumb people can easily communicate with other normal people. The system can be used by the user without any difficulty and complexity. The application is cost efficient and eliminates the usage of expensive technology.

ACKNOWLEDGMENT

The author would like to express my gratitude to all colleagues at Mandalay Technological University who have contributed to the preparation of this research work.

REFERENCES

[1]. Shweta Sonajirao Shinde, Dr. R.M. Autee, Real Time Hand Gesture Recognition And Voice Conversion System For Deaf And Dumb Person Based On Image Processing, IJRPET, *ISSN: 2454-7875, VOLUME 2, ISSUE 9, September -2016.*

[2]. Guat Yew Tan1*, "A Hybrid Convex Hull Algorithm for Fingertips Detection", Indian Journal of Science and Technology, Vol 9(28), DOI: 10.17485/ijst/2016/v9i28/97821, July 2016.

[3]. Ying Wu and Thomas S. Huang, "Hand modeling, analysis and recognition For Vision-Based Human Computer Interaction", IEEE Signal Processing Mag, p.51-60, May 2001.

[4]. Japneet Kaur , Ramanpreet Kaur, "Color Image Reduction Techniques and Optimization of Penalty Functions ", International Journal of Scientific Research Engineering & Technology(IJSRET) Volume2, Issue3, pp 158-162, June 2013.

[5]. S. J. Wan, P. Prusinkiewicz, and S. K. M. Wong, "Variance based color image quantization for frame buffer display," Color Res.Applicat., vol.15, no.1, pp.52–58, 1990.