INDIAN SIGN LANGUAGE TO SPEECH

Kusumika Krori Dutta, B.Sunny Arokia Swamy,
Anil Kumar G S, Konduru Satheesh Kumar Raju
Dept. of Electrical and Electronics Engineering , M.S.Ramaiah Institute of Technology
kusumika@msrit.edu sunny23june@gmail.com anilk6871@gmail.com satheeshkumarraju21@gmail.com

ABSTRACT

Communication is a key factor and it is essential for good human relations. Over the ages, sign language has progressed and now in certain countries, it has developed so much as to acquire status equal to the spoken and written language. Here, Indian sign language is converted to speech using image processing. Images from a high definition digital camera are processed and the same is converted to Text after fixing threshold to image. The text is later converted into speech. Two different algorithms has been used for feature extraction. One algorithm is based on parameters and other is by using centroid algorithm.

Index Terms— Indian Sign Language, Image Processing, Digital camera, Lab color Space, Centroid.

INTRODUCTION

Communication is the way to express the feelings and exchange information. Without this basic means, human life would have been non-existent. But normal (oral) communication is not possible for the hearing impaired and mute people. Over 5% of the world's population i.e., 360 million people – have disabling hearing loss (328 million adults and 32 million children). The majority of these people live in low- and middle-income countries. In India, approximately 63 million people (6.3%) are suffering from significant auditory loss.

In order to overcome this problem sign language was introduced. About the origin of sign language, nothing much is known. The earliest references are about a Christian sect whose members took vow of silence and communicated with each other through signs and gestures. Over the ages, sign language has progressed and now in certain countries, it has developed so much as to acquire status equal to the spoken and written language. Most of the nations in the world have their own sign language. Because of the fact that some sort of silent mode of communication has been there all along and the fact that a child born deaf, learns to express him/herself through signs, sign language came to be regarded as the mother tongue of the deaf people.

Sign language is a language which uses gestures to convey messages. It is the orientation and movement of hand, arms and body combined with facial expressions. Though sign language is the best known solution for communication of this segment of disabled people, the interaction using this language is restricted to people having the knowledge of it and need to depend on a translator

in case of interaction with rest of the world [1]. This eventually reduces opportunities of disabled people and act as an impediment towards natural interaction process.

The system presented in this paper provides an opportunity for hearing impaired and mute people to communicate in general without any translator. In this paper, a system is designed and developed to convert gestures made in sign language to speech. A digi-cam is interfaced with PC via MATLAB. An image of the sign language was captured with the help of digi-cam [2]. Later a particular threshold is fixed based on ten different images taken for a particular gesture. Then using the conditional statement the gesture was properly recognized and converted to text. This text was further processed to speech.

This vision based system is compact and feasible compared to the mostly available glove based sign language translator system [3-4]. Also regardless of knowledge about the system, it can be accessible universally since the output of system is speech. Also there is a display of message in the form of text which can be used directly to compose electronic mails and messages without any additional composing. This provides greater advantage for ambitious illiterate disabled people who faces difficulty in doing things. We can even make the speech in different local languages which is more beneficial to that particular region.

PROPOSED METHODOLOGY

The images are acquired from the camera, processed and corresponding speech output is given to the speaker [5-6]. For feature extraction two algorithms were used one was based on different parameters like area, entropy and so on. While other algorithm was based on considering centroid of the segmented region.

The proposed system can be divided into three stages:

A. Image Acquisition
B. Segmentation
C. Feature extraction
    I) Statistical Method
    II) Centroid Algorithm
D. Speech Output

A. Image Acquisition

Image Acquisition was done using a higher resolution digital camera.The digi-cam was interfaced with PC installed with MATLAB via USB-port. At a given point of time, five frames per trigger of a single image was initially captured and this acquired image was processed. The acquired image representing number '1' (one) can be seen in Fig 1. Similarly other numbers from '2' (two) to '5' (five) were captured.

Fig 1: Image representing the number '1'(one)

B. Image Processing

The acquired image was in the RGB format. As a pre-processing technique, Color transformations RGB to L*a*b* was done with the acquired image. Consider the Fig 1 showing the number '1' (one). This RGB image was transformed into Lab color space which is a color opponent space with dimension L* for lightness while a* and b* for the color-opponent dimension. The Lab image is shown in Fig 2. The corresponding equations for calculating the dimensions (L*, a*, b*) was indicated below Eq. 1, Eq. 2, Eq.3. Second layer of lab color space a* is shown in Fig 3 while third layer b* is shown in Fig 4.

$$L^* = 116\, f(Y/Y_n) - 16 \quad (1)$$

$$a^* = 500[f(X/X_n) - f(Y/Y_n)] \quad (2)$$

$$b^* = 200[f(Y/Y_n) - f(Z/Z_n)] \quad (3)$$

Where,

$$f(t) = \begin{cases} t^{1/3} & \text{if } t > (6/29)^3 \\ 1/3\,(29/6)^2\, t + (4/29) & \text{otherwise} \end{cases}$$

Where X_n, Y_n and Z_n are the CIE XYZ tristimulus values of the reference white point.
Further, the gray threshold was found for the second layer and third layers i.e., for a* and b*. Those layers were converted to binary image. The binary image for second layer is shown in Fig 5 and binary image for third layer is shown in Fig 6. These two binary images of a* and b* were multiplied pixel by pixel and the resultant binary image was obtained. Fig 7 is the final binary output for number '1'(one). Figures 12 to 15 represent acquired and binary output images of different numbers

C. Pattern Recognition or Gesture Recognition

*I) Statistical Method***:**

In order to identify the gesture, few parameters were calculated like Area, Entropy, perimeter, Row width, Column width, form factor, centroid, 1st order differentiation on column wise, 1st order differentiation on row wise and Eccentric ratio . These parameters were found to be different for each and every type of gesture. Area is the extent of a two-dimensional surface or shape, or planar lamina, in the plane. It can also be defined as amount of flat space that it covers.

In the case of an image, entropy corresponds to the number gray levels which the individual pixels can adopt. Eccentric ratio is a parameter associated with every conic section and it is the measure of how much the conic section deviates from being circular.

Here row width and column width were calculated to measure Eccentric Ratio using equation Eq. 4.

$$\text{Eccentric Ratio} = C_{max} / R_{max} \qquad (4)$$

Form Factor was calculated using the Eq. 5 using area and perimeter.

$$\text{Form Factor} = (4\pi * \text{area}) / \text{perimeter}^2 \qquad (5)$$

Single gesture with 10 different images were acquired and the proper threshold was fixed. After fixing the threshold using the conditional statements the gesture was properly recognized and converted into text.

II) Centroid Algorithm: In this method, Centroid or Centre of gravity of the hand region in binary output image was found using the mathematical equation Eq. 6. Fig 8 showing the binary image representing the number '5' (five) was considered as a sample.

$$x' = \frac{\sum_{i=0}^{k} xi}{k} \qquad y' = \frac{\sum_{i=0}^{k} yi}{k} \qquad (6)$$

Where x_i and yi are (x, y) of the i^{th} pixel in the hand region, while k denotes the number of pixels in particular given region.

Fig 9 shows the binary image with centroid mark. After obtaining the centroid of hand, the distance of farthest point in the hand region from the centroid was found [7]. Two concentric circular windows one with radius 0.9 times other with radius 0.8 times the length of farthest point

respectively was formed and these two windows were multiplied with obtained binary image. Fig 10 and Fig 11 are two circular windows. The number of fin ger were found from the resultant image and it was displayed as text.

D. Speech Output:

The text was further processed using TTS algorithms in MATLAB. Once the text is obtained, it is converted into speech such that voice is obtained as an output. This was done using 'actxserver' in the MATLAB which creates a custom interface server on a remote machine. This ensures that even the illiterate people can understand the message more easily and effectively.

RESULTS AND DISCUSSION

A. Statistical Method

From Table 1, we can observe that area for number '1' (one) was 24152 later it was increased as we move from top to bottom of the table finally this increment ends up at 41044 for number '5'(five). Similarly Perimeter also increases as we move from top to bottom in table from 791 for number '1' to 1598 for number '5'. Entropy also increases from 0.8983 for number '1' to 0.9966 for number '5'. And Form factor decreases from 0.4851 to 0.202. First differential column wise increases from 275 for number'1' to 677 for number '5' similarly for first order differential row wise increases from 656 for number '1' to 1414 for number '5'.

The results were obtained by making assumptions that background of the hand should be plain for image acquisition and the distance between the digi-cam and hand should remain constant.

B. Centroid Algorithm

For one of the sign, number '5' (five) the algorithm was applied and active region was determined. The same algorithm can be utilized and applied for different types of signs. Since this algorithm is invariant to rotation, translation and scale of the hand, this is best for different types of people irrespective of age, color and position of hand in given region. This algorithm is best suited for plain background but not for cluttered image.

CONCLUSION

This project will help and aid the hearing impaired and mute people to progress from dependent to independent life. It will give them the confidence and also they can comfortably share their thoughts, experience, ideas, and difficulties to other segments of the society and will be able to live a normal life.

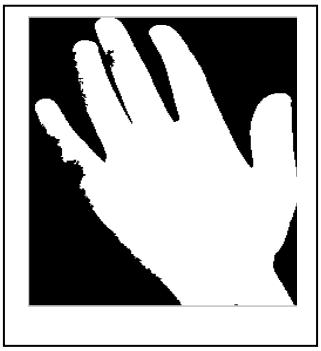

Fig 8:Final binary output '5'

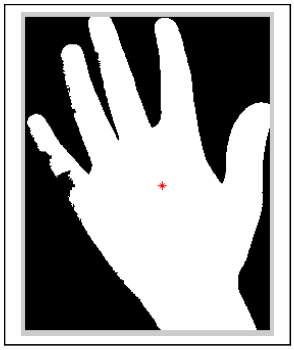

Fig 9: Binary image with centroid

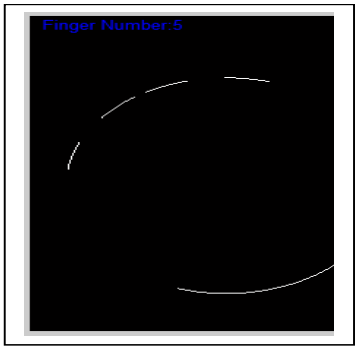

Fig 10:First circular window

Fig 11:Second circular window

Fig 12: Image representing '2'(two)

Fig 13: Image representing '3'(three)

Fig 14: Final Binary image Representing '2'(two)

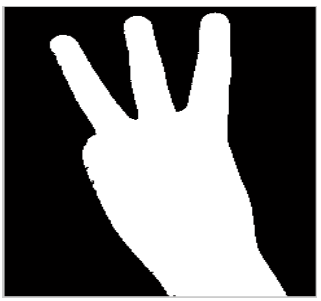

Fig 15: Final Binary image Representing '3'(three)

REFERENCES

[1] Praveenkumar S Havalagi, Shruthi Urf Nivedita, "The amazing digital gloves that give voice to the voiceless", International Journal of Advances in Engineering & Technology, vol 6, Issue 1, pp. 471-480, Mar. 2013.

[2] M Jerin Jose, V Priyadharshini, M Suresh Anand, Dr. N. Mohan Kumar, " Indian sign language (ISL) translation system for sign language learning", International journel of innovative research and development, vol 2 Issue 5, May 2013.

[3] Niklesh Thakre, Nitin Shahare, Anil Balbudhe, "Voice synthesizer for deaf and dumb using ARM", Discovery, Volume 19, Number 56, May 12, 2014.

[4] Gunashekaran K, Manikandan R, "Sign language to speech translation system using PIC microcontroller", International Journal of Engineering and Technology (IJET), vol 2, Number 2 , Apr-May 2013.

[5] Ashish Sethi, Hemanth S, Kuldeep Kumar, Bhaskara Rao N, Krishnan R, "Signpro - An application suite for deaf and dumb", IJCSET, Vol 2, Issue 5, May 2012.

[6] Rafiqul Zaman Khanand, Noor Adnan Ibraheem, "Hand gesture recognition: A literature review", International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012.

[7] W. T. Freeman and M. Roth, "Orientation Histograms for Hand Gesture Recognition", IEEE International Conf. on Automatic Face and Gesture Recognition, 1995.