Sign Language Recognition for the mute people

Subham Agarwal
School of Computing Science and Engineering
Vellore Institute of Technology
Chennai, India 9038203497
mr.subham.agarwal@gmail.com

Komal Khetlani
School of Computing Science and Engineering
Vellore Institute of Technology
Chennai, India 8987272601
komal.khetlani2525@gmail.com

***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 or a predefined phrase. The system identifies the hand from live video and uses image processing techniques to identify the gesture. The basic objective of this project is to develop a computer based intelligent system that will enable mute 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 and artificial intelligence concepts to take visual inputs of sign languages 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.**

I. Introduction

Mute 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.As like any other language it has also got grammar and vocabulary but uses visual modality for exchanging information. The problem arises when mute 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 alike people.At this age of Technology the demand for a computer based system is highly demanding for the mute community.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).Gestures are the non-verbally exchanged information. A person can perform innumerable gestures at a time.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 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 1970s. 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 systems 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

The project consists of two sectionsc

A. Character Recognition

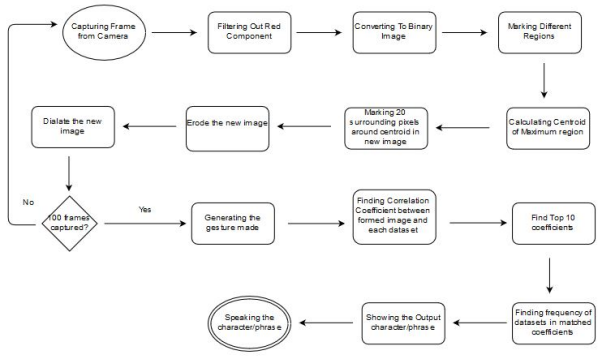

Fig. 1. Flowchart

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.

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.

2) 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).

3) 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.

4) Identifying the Finger and making the gesture: To identify the image the user sticks a red tape or any red object to his/her fingers. From the captured frame we slice only the red plane and subtract it from the grayscaled image to get only the red components in the image.Since the project requires no red objects to be present in the background so the red objects are identified as fingers in the image.Now the centroids of the red regions are calculated.In a new blank white image the centroids of all the red objects are marked as black along with 20 surrounding pixels(for creating visaully better images).

5) Processing the image: The image formed consists of many irregularities and sharp edges . The image is subjected to many filters to make it more recognizable. The following filters are applied one after the other in succession.

1) Median filter : This filter is applied to remove any noise present in the image and to blur and smoothen the edges present in the image
2) Erosion filter : Erosion removes small-scale details from a binary image but simultaneously reduces the size of regions of interest, too. By subtracting the eroded image from the original image, boundaries of each region can be found
3) Dialation filter : The dilation of an image f by a structuring element produces a new binary image Dilation has the opposite effect to erosion it adds a layer of pixels to both the inner and outer boundaries of regions.

6) Recognizing the gesture: Recognition of the image requires it to be compared to different datasets and the most matching one is shown as the output. For the project we have 50 images for 5 characters (10 images for each character in the dataset) .Correlation coefficient are calculated between the made gesture and all the datasets and storing the result of each comaparision in an array.After comparing with all 50 dataset images the 10 maximum coefficients are selected , the 10 selected coefficients are now counted for to which of the 5 dataset images do they map to .The frequency of occurance of each image is stored in a new array and the maximum frequency count image is displayed as the most matching character/ phrase.

B. Hand Gesture Recognition

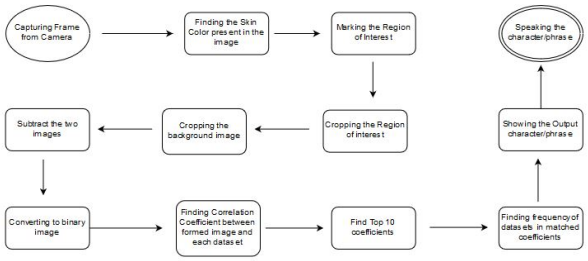

Fig. 2. Flowchart

Fig 2 shows the flowchart for implementation of Hand Gestures. The different steps are described in detail in the following sections.The image is first acquired using the webcam.Different filters are applied on the image to remove noise.The image is then converted to ycrcb format.The image is thresholded to identify only the skin colors present in the image.The largest blob of skin detected is segmented and is identified to be the hand.The image is pre processed for better calculation of correlation coefficient.Correlation coefficient is calculated between the formed iamge and all the dataset images to find the most matching one.The most matched dataset is shown as output and the particular phrase is read.

1) Image Acquisition : The image is acquired using a webcam and is flipped to make the orientation as per the user. From the camera images are captured in frames.

2) Region of Interest Identification: At the start the first frame is taken and saved as background image. On each input frame median filter is applied to remove the noise.Then the image is converted to YCbCr format for easy processing. The image is then thresholded in the pixel region 70-120 in Cb and 80-143 in Cr range which represents the skin color pixel values.

3) Region of Interest Segmentation: All the different regions in the thresholded are marked and the centroid,area of all these regions are calculated .The region with the maximum area becomes are area of interest and 256,256 pixels fromthe centroid of the region are cropped out of the image. The same region is cropped from the background image and both the image are subtracted and then thresholded to form the gesture.

4) 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]

5) Thresholding : Thresholding is the simple method of image segmentation[1]. In this method we convert the RGB image to Binary image. . 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.

6) Recognizing the gesture: Recognition of the image requires it to be compared to different datasets and the most matching one is shown as the output. For the project we have 50 images for 5 gestures (10 for each gesture) .Correlation coefficient are calculated between the made gesture and all the datasets and storing the result of each comaparision in an array.After comparing with all 50 dataset images the 10 maximum coefficients are selected , the 10 selected coefficients are now counted for to which of the 5 dataset images do they map to .The frequency of occurance of each image is stored in a new array and the maximum frequency count image is displayed as the most matching character/ phrase.

IV. Conclusion and Future Work

The project stands as a intermediatary mode of communication between the mute and the rest of the world. It makes the mute people overcome their problem of communicating and expressing their ideas to the world. A boundary-trace based finger detection technique is presented and cusp detection analysis is done to locate the finger tip. This algorithm designed is a simple, efficient and robust method to locate finger tips and enables us to identify a class of hand gestures belonging to the American Sign Language which have fingers open. The accuracy obtained in this work is sufficient for the purposes of converting sign language to text and speech since a dictionary can be used to correct any spelling errors resulting from the 5% error in our gesture recognition algorithm. In future work, sensor based contour analysis can be employed to detect which fingers in particular are open. This will give more flexibility to interpret the gestures. Furthermore, hand detection method using texture and shape information can be used to maximize the accuracy of detection in cluttered background. More importantly, we need to develop algorithms to cover other signs in the American Sign Language that have all the fingers closed. An even bigger challenge will be to recognize signs that involve motion (i.e, where various parts of the hand move in specific ways).

Acknowledgment

The authors would like to thank Dr. Geetha S., Faculty School Of Computing Science And Engineering VIT University and Prof. Priyadarshini J., Faculty School Of Computing Science And Engineering VIT University for their continued support and inputs throught the completion of the project.

REFERENCES

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.

[2] Y. Shirai, N. Tanibata, N. Shimada, Extraction of hand features for recognition of sign language words, VI'2002,Computer-Controlled Mechanical Systems, Graduate School of Engineering, Osaka University, 2002.

[3] C. Nlker, H. Ritter, Detection of Fingertips in Human Hand Movement Sequences, Gesture and Sign Language in Human-Computer Interaction, I. Wachsmuth and M. Frohlich, eds., pp. 209-218, 1997.

[4] . Bauer and H. Hienz, Relevant features for video-based continuous sign language recognition, in Proc. of Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 440-445, March 2000.

[5] . Hamada, N. Shimada and Y. Shirai, Hand Shape Estimation under Complex Backgrounds for Sign Language Recognition , in Proc. of 6th Int. Conf. on Automatic Face and Gesture Recognition, pp. 589-594, May 2004. Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, March 18 - 20, 2009, Hong Kong