Gesture based text recognition and text to speech conversion

Member:K Upendra Sainath Reddy,201601038,cse

Abstract—**This System is a prototype system that helps to recognize hand gesture to normal people in order to communicate more effectively with the special people. This focuses on the problem of gesture recognition in real time that sign language used by the community of deaf people. The problem addressed using Skin Detection, Image Segmentation, Image Filtering, and Template Matching techniques. This system recognizes gestures of ASL (American Sign Language) including the alphabet.**

I. INTRODUCTION

Communication means to share thoughts, messages, knowledge or any information. Since ages communication is thetool of exchange of information through oral,writing, visuals signs or behavior. The communication cycle consider to be completed once the message is received by a receiver and recognizes the message of the sender. Ordinary people communicate their thoughts through speech to others.

whereas the hearing impaired community the means of communication is the use of sign language and ASL is 3rd most used sign Language.

II. OBJECTIVE

This system focuses on the problem of gesture recognition in real time that sign language used by the community of deaf people.And uses Color Segmentation,Skin Detection, Image Segmentation, ImageFiltering, and Template Matching techniques like Correlation.

This system recognizes gestures of ASL including the alphabet.And uses Template matching for matching the gesture to keep it simple,other methods include neural networks.

III. What is ASL?

ASL (American Sign Language) is a language for hearing impaired and the deaf alike people, in which manual communication with the help of hands, facial expression and body language are used to convey thought to others without using sound. Since ASL uses an entirely different grammar and vocabulary, such as tense and articles, does not use "the", therefore, it is considered not related to English. ASL is generally preferred as the communication tool for deaf and dumb people.

IV.Concerns and Limitations

Visibility issue may arise due to several reasons For instance, the camera where the user has to stay in position, the various environmental conditions like lighting sensitivity, background color and condition, electric or magnetic fields or any other disturbance may affect the performance.

- The sitting or standing position of the signer may vary in front of camera. Movements of the signer, like rotating around the body must be taken into account.
- Delay in the processing execution can be occurred due to the large amount or higher resolution of image. Hence, it is difficult to recognize in real time basis.

V. Image Acquistion

The common method of image acquisition can be done from digital photography usually include Digital Camera.A simple Gui is used to take frames

from live video and to detect Guestures made by the signer.

Figure 1: Gui to take input video

VI. Image Processing Steps

To satisfy and reduce the computational effort needed for the processing, pre-processing of the image taken from the camera is highly important. Apart from that, numerous factors such as lights, environment, background of the image, hand and body position and orientation of the signer, parameters and focus the of camera impact the result dramatically.

VII. Skin Detection

There are several techniques used for color space transformation for skin detection. Some potential color spaces that are considerable for skin detection process are:

- CIEXYZ
- YCbCr
- YIQ
- YUV

And YCbCr metric is used for Skin Detection algorithm. Skin Detection process involves classification of each pixel of the image to identify as part of human skin or not by applying Gray-world Algorithm for illumination compensation and the pixels are categorized based on an explicit relationship between the color components YCbCr. In YCbCr colorspace, the single component "Y" represents luminance information, and Cb and Cr represent color information to store two color-difference components, Component Cb is the difference between the blue component and a reference value, whereas component Cr is the difference between the red component and a reference value.

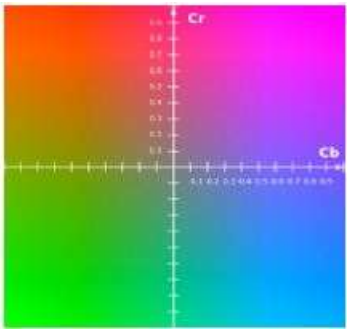


Figure 2 : Cb and Cr color composition

Gray world algorithm

Gray world algorithm is used for the input picture taken from video for illuminance compensation. The gray world normalization makes the assumption that changes in the lighting spectrum can be modelled by three constant factors applied to the red, green and blue channels of color.By this we can achieve image normalization in the input picture Which helps us to simplify and increase the ability of separation between skin and non-skin, and also decrease the ability of separation among skin tone.

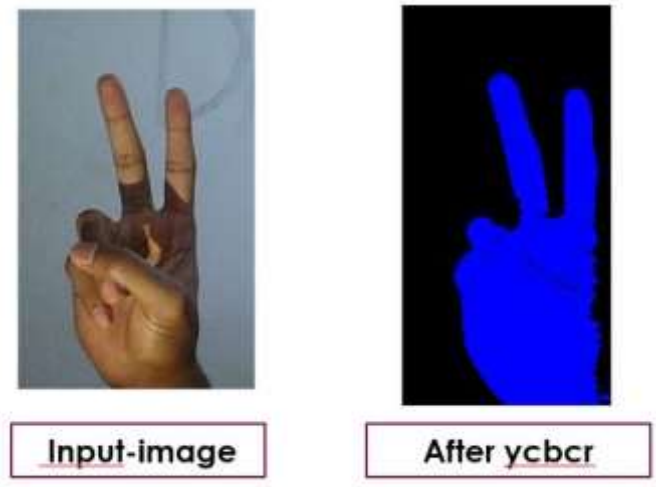


Figure 3 : Skin pixel Detection

Thus, a pixel is considered a human skin, if a set of pixel is falling into that particular category with a certain value of Cr and Cb having certain threshold.
Thus Condition for a skin pixel is

$$skin=\begin{cases}1 \; if (Cb \geq 77 \; and \; Cb \leq 127 \; and \; Cr \geq 133 \; and \; Cr \leq 173) \\ 0 \; otherwise \end{cases}$$

And the detected skin pixels are marked as blue for easy detection. After the skin detection, image marked with Blue color converted into the binary with skin pixels as '1' and rest are "0". So that, the correlation of the image can be matched with the Template.

VIII. Image Filtering

Image Filtering is applied for the image to remove noise in the gesture.And median Filter is applied to filter the image.After applying median filter each output pixel contains the median value in a 3-by-3 neighborhood around the corresponding pixel in the input image. Median filter pads the image with 0s on the edges, so the median values for points within one-half the width of the neighborhood ([m n]/2) of the edges might appear distorted.

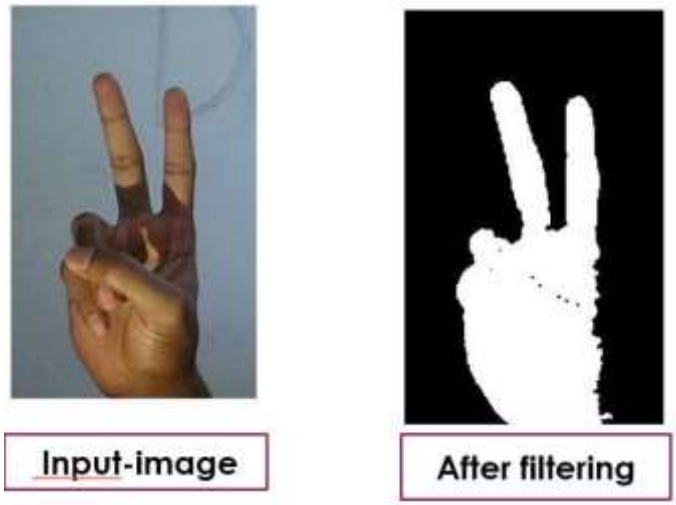


Figure 4: After applying median filter

IX. Template Matching

Template matching is done by using correlation. It involves determining correlation coefficient between two image one is template image and another is search image. Template matching involves an predefined gesture database which is used to match with the input gesture.
The input gesture is matched with every predefined image gesture in database and it's corresponding correlation coefficient is determined and the image with highest correlation coefficient is determined as matched gesture and it's corresponding alphabet is determined. And the determined alphabet is concatenated to a string and this string is used for speech conversion.
Correlation coefficient is a statistical measure of the degree to which changes to the value of one variable predict change to the value of another. And it is Calculated by the formula

$$r = \frac{\sum_m \sum_n (A_{mn} - \overline{A})(B_{mn} - \overline{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \overline{A})^2\right)\left(\sum_m \sum_n (B_{mn} - \overline{B})^2\right)}}$$

Correlation coefficients are expressed as values between +1 and -1. A coefficient of +1 indicates a perfect positive correlation, A change in the value of one variable will predict a change in the same direction in the second variable. A coefficient of -1 indicates a perfect negative correlation, A change in the value of one variable predicts a change in the opposite direction in the second variable. Lesser degrees of correlation are expressed as non-zero decimals. A coefficient of zero indicates there is no discernable relationship between fluctuations of the variables.

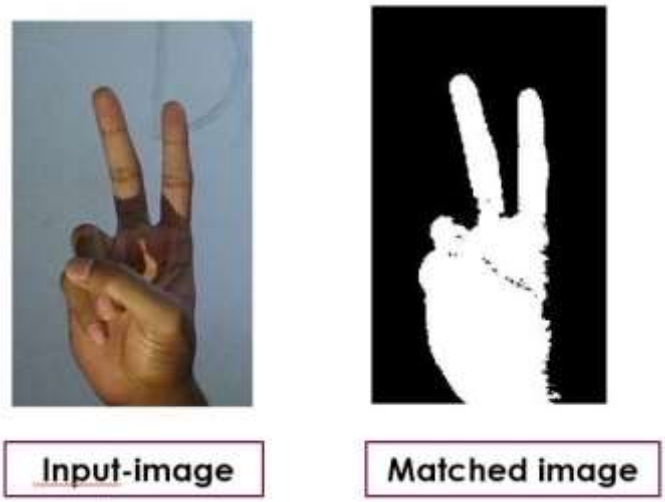


Figure 5 :Template Matching

The matched image is then used to determine the corresponding alphabet in ASL Language.

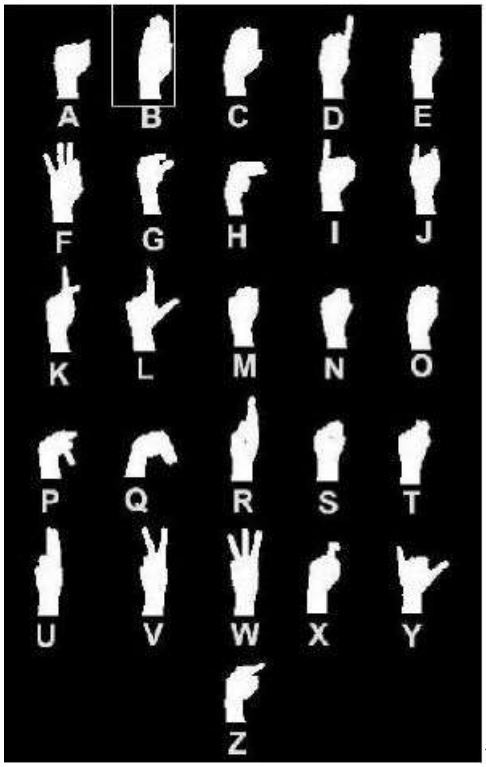


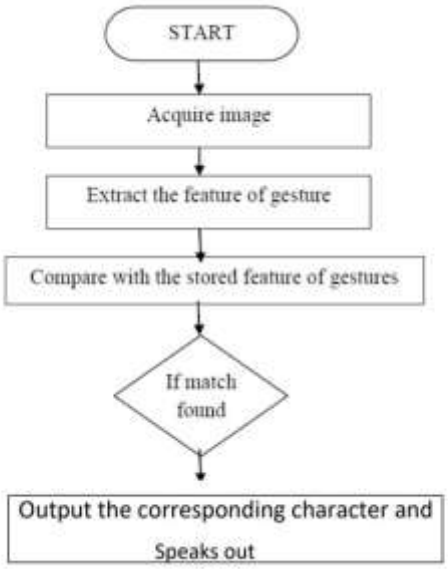
Figure 6: American Sign Language(ASL)

X. String Conversion and speech conversion

The determined alphabet from ASL for the corresponding gesture is concatenatd to a string so that it can used for the speech conversion for which this stiring serves as an input .And the string is used to speak out the word using speech function.

XI. Implementation

This system implements the process of detection of handgesture and speech conversion in the following Order.



XI. Result And Analysis

The purpose of this application is to recognize hand gesture. The design is very simple and the signer doesn't need to wear any type of hand gloves

DISTANCE

- This system uses a webcam to capture gestures and the distance that it can detect an human hand is upto 100cms and may vary in other devices
- The system may malfunction if the signer is in crowded place as it may detect more skin pixels other than the signer
-

TIME

- The captured gesture requires approximately 2 seconds to match an gesture from the database for the input sign
- And it involves looping through all the predefined gestures in databse and computing the correlation coefficients and determining the gesture with high correlation coefficient
- And the timer is set to 30 seconds to extract the hand gesture from the video input.

This system can recognize a set of 24 letters from the ASL alphabets: A, B, C, D, E, F, G, H, I, K, L, M, N, O, P, Q, R, S, T, U, V, W.

XII. CONCLUSION

This system can be used in mobile devices for the ease of use and the accuracy of the gesture matching can be increased by training the dataset than template matching using correlation.
Overall the sytem works perfectly for some gestures and doesn't need any other requirements to attain accuracy.The image acquisition and skin detection works with high accuracy.The results obtained are applicable,and can be implemented in a mobile device smart phone having frontal camera.

XII. REFERENCES

[1] Yang quan, "Chinese Sign Language Recognition Based on Video Swquence Appearance Modeling", ICIEA, the 5th IEEE Conference, pp: 1537 – 1542, 2010

[2] http://web.stanford.edu/~sujason/ColorBalancing/grayworld.html

[3] http://in.mathworks.com/help/images/ref/bwboundaries.html