Hand Gesture Recognition And Conversion to Text & Speech

Presented By :

Meghna Sinha - 102196002

Devansh Kaushik - 102066005

Bhoomica Gupta - 102016070

Utkarsh Chauhan - 102066006

Team - 23

CONTENTS

OBJECTIVE

The techniques available in our field require use of gloves with sensors, pointers, heavy GPUs or a specific background with any disturbance or noise making them expensive and impractical.

In our approach we aim at extracting features of hands using OpenCV functions for converting hand gestures to alphabets as per the American Sign Language.

SYSTEM ARCHITECTURE

STEP 1

- Input captured with camera
- Noise removal & smoothening

STEP 2

- Input converted to grey scale
- Otsu method for thresholding is done

STEP 3

- Contour extraction

STEP 4

- Depending on the number of convexity defects, convexity hull or haar cascade classifier is applied

STEP 5

- Alphabet Identified and Displayed

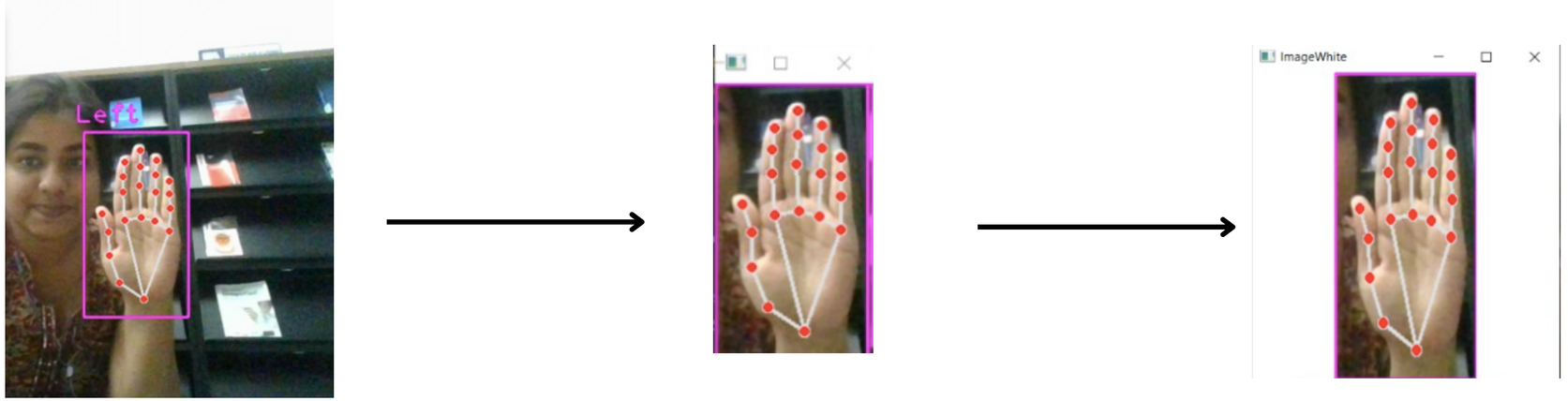
CAMERA MODULE

This module is responsible for connecting and capturing input through the different types of image detectors and sends this image to the detection module for processing in the form of frames. Data gloves, hand belts and cameras are commonly used for image capture.

Our technique require the use of web cam, which is cost efficient and easily available

DETECTION MODULE

- Detection module is mainly responsible for detection of object, in this case 'The Hand'. Colour conversion, noise removal and image smoothening also takes place in this module.
- The image is first cropped to a size 300 * 300 pixels (since the whole image needs to be of the hand)
- The resized image is then centred against a white background, since the images need to have same sizes and in ASL hand gesture vary in their height and width.



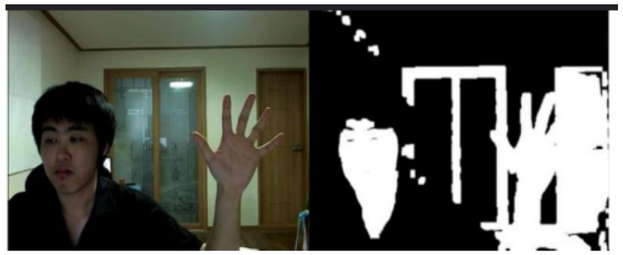
Hand detected Cropped Image Centred against a background

Implementation of detection module

THRESHOLDING

- The cropped image is turned into grey scale image, which is then converted to a binary image using thresholding.
- RGB colour space, YCbCr thresholding, inverted binary thresholding and Otsu's thresholding are among the popular methods.
- Otsu's method has been applied in our model.
- Otsu's method aims at minimising intra-class variance and maximising inter class variance.

RGB Color space with threshold value 70



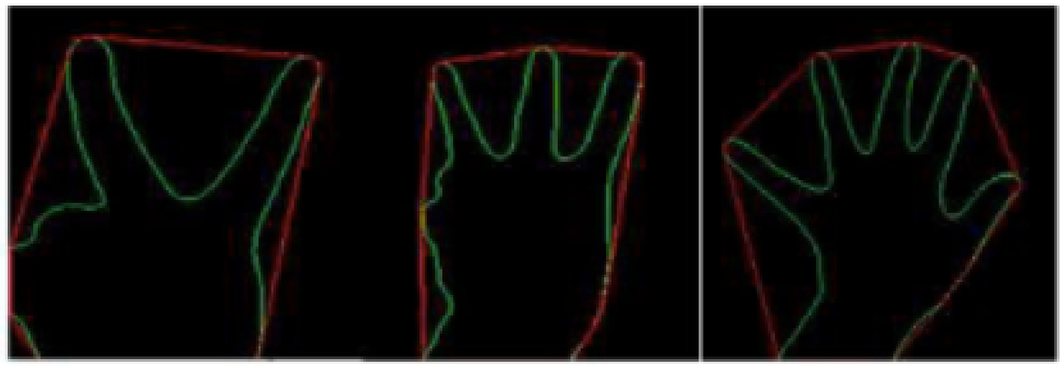
YCbCr threshold image

Result of Otsu's method

Side by side comparison of various thresholding techniques

CONTOUR EXTRACTION

- Contour analysis is most useful when it comes to object detection.
- Contours are curves that link continuous points of same colour.
- Once contouring is done, convexity hull or haar cascade is applied depending upon the number of defects present



Contour Extraction of hand gestures

GESTURE IDENTIFICATION

- ***Convexity Hull and Convexity Defects***

  Convex hull for the hand will be the convex polygon which is surrounded by the vertices corresponding to the tip of the fingers. Convexity defect is the calculated difference between the convex hull and the contour. This method determines the number of fingers in the gesture and then uses it to identify the gesture.

- ***Haar Cascade Classifier***

  Haar Cascade Classifier is used in case of no convexity defects. Haar Cascade Classifier requires a minimum of 10 positive images in different orientation as well as lighting conditions and a minimum of 1000 negative images for proper training.

RESULTS

Alphabet A Identified

Alphabet B Identified

Alphabet C Identified

Efficiency :

- Average efficiency of 92.6% achieved with plain background
- Average efficiency is only 70% in a non-plain background

SPEECH CONVERSION OF THE IDENTIFIED LETTER

- Once the letter is identified the audio part is handled by bash and shell script
- On identification of letter a directory is made.
- Then a .txt file containing the identified letter is made. If the file already exists then it is overwritten.
- Shell Script then identifies the latest modified file and plays the corresponding audio file.
- This ensures that the sound is played continuously and without disturbance.

CONCLUSION

Our project aims for a robust recognition system that will not utilize any markers, hence making it more user friendly and low cost. In this system we aim to provide all the 26 alphabets as per ASL.

However our project is limited to only single hand gestures. In the future we would like to extend our project to both hands in order to formulate letters and also aim at improving accuracy of our current model.