Real Time Hand Gesture Recognition Using Different Algorithm Based On American Sign Language

Prof. B.B.Gite

bbgite.sae@sinhgad.edu

Vaibhavi Honmane

vaibhavihonmane2015@gmail.com

Vaibhavi Joshi

Vmj31898@gmail.com

Rutuja Waghe

rutuja.waghe@gmail.com

Sinhgad Academy of Engineering, Kondhwa (Bk.), Pune

Savitribai Phule Pune University, Pune

Abstract

Sign Language research field based on real time hand gestures called sign samples and recognition unit of computer. The sign language is the only way to be used for deaf and dumb community communication platform. In this system, we are working on the American Sign Language(ASL) dataset (A-Z) alphabet recognition followed by our word recognition dataset of Indian Sign Language (ISL). Sign data samples to be making our system more accurate with help of Convolutional Neural Network (CNN). Today, much research has been going on the field of sign language recognition but existing study failed to develop trust full communication interpreter. The purpose of this system is to represent a real time two way communication interpreter based on Indian Sign Language (ISL) with higher accuracy. Indian Sign Language (ISL) used by Deaf people community in India, does have acceptable, meaningful essential and structural properties.

Keywords - Artificial Neural Network, Indian Sign Language, Hand Gesture Recognition, Deaf community.

I. INTRODUCTION

There are so many languages in India used as officially and locally. Such large diversity country has more challenges to maintain uniqueness in language interpretation. In languages have its challenges when it used to communicating over different areas, societies and states. Indian Sign Language (ISL) is one of the living languages in India used by the Deaf community peoples. But as we seen there is not any standard language available till date. So we are working on different sign language dataset to invent Indian sign language as an interpreter.

We are going to implement two-way communications for Sign language is used for the people who are deaf or hard of hearing and also used by them who can hear but cannot physically speak. Our motive behind this implementation is to create complete language which involves movement of hands, facial expressions and gesture of the body. The Sign language is not universal standard so we are making our contribution towards sign language development. Every country has its own native sign language like

American Sign Language work for alphabet recognizer. Each sign language has its own rule and semantic meanings. The problem comes when deaf and dumb people want to communicate or trying to say something there is not any language for them. So it becomes necessary to develop an automatic language interpreter to assist them for their fluent communication. They people want something more helpful which makes their communication universal and easy. Another one is based on computer vision-based gesture recognition, which involves image processing techniques. Consequently, this category faces more complexity. Our motive to develop this system based on real time signs.

This system captures hand gesture images of ISL with system camera for feature extraction. The analyzing phase, pre-processing unit is used to the noise removal, grey scale conversion by using Gaussian filter, binary conversion of images done by using OTSU's method followed by feature extraction. In our system, Convolutional Neural Network (CNN) is going to be used for future recognition in which we having the input unit of training data set of images. Next we have hidden unit which acts upon this training dataset to evaluate the output unit results train model. This entire CNN works by considering the factors namely matrix feature of images for drafting into a train model for real time sign recognition. The working with real time sign language we know that the dataset need to be large and rich in processed features.

II. METHODOLOGY

Open-CV: Open-CV was designed for computational efficiency with a strong focus on real time applications. Open-CV was built to provide a common infrastructure for computer vision application and to accelerate use of machine perception in commercial product. Open-CV makes it easy for businesses to utilize and modify the code.

Gaussian filter: Gaussian filter is a linear filter used to blur the image and to reduce noise. It alone will blur edges and reduce contrast. Gaussian filter modifies the input signals by convolution with a Gaussian function. In electronics and signal processing, a Gaussian filter is a filter whose impulse response is a Gaussian function.

It is used in numerous research areas like it defines a probability distribution for noise or data also used in mathematics and it is smoothing operator. Gaussian filter is non-uniform low pass filter, it might not pressure image brightness

Color to gray-scale conversion: In color image each pixel is represented by a triple(R, G, B) i.e. intensities for colors Red, Green, Blue. The gray-scale image is represented by 8 bits value of luminance. In gray-scale image luminance of pixel value ranges from 0 to 255. Converting color image to gray-scale is nothing but converting RGB values (24 bit) to gray-scale value (8 bit).

Conversion of color image to gray-scale image is done using Open-CV. In image software there are three algorithms for this. First is lightness method, it takes the average of most and least prominent colors. Second is average method, this method computes average of R, G, B values of image. And the third is luminosity method it is one of the complex version of average method. This method also takes average but it derives weighted average to account for human perception. Humans are more sensitive to green color so green is weighted heavily.

Gray to Binary: The simplest and most common way of image segmentation and convert gray-scale image to binary image is thresholding. The thresholding includes, selection of thresholding value (T) and then all the gray value below threshold value are classified as 0(black i.e .background) and all the gray level value which is equal to or greater than threshold value are classified as 1(white i.e. foreground).

$$g(x, y) = 1 \;;\; \text{if } f(x, y) >= T$$

$$0 \;;\; \text{otherwise}$$

Where,

$g(x, y)$ = threshold image pixel at (x, y)

and $f(x, y)$ = gray-scale image pixel at (x, y)

imbinarize method is used to convert gray-scale image to binary image by replacing all pixels in the input image with luminance greater than determined threshold with 1 and setting other values 0.

BW = imbinarize(I) converts the 2-D or 3-D gray-scale image I to binary image BW.

CNN: CNN is nothing but Convolutional Neural Network. CNN is just kind of ANN(Artificial Neural Network). CNN has one or more layers of convolution units. The major difference between CNN and ANN is in ANN each neuron is connected to every other neuron whereas in ANN only last layer is fully connected.

In proposed system CNN is used to recognize the input hand gesture. There are four layers in CNN as convolution layer, pooling layer, flattening layer and fully connected layer. Feature extraction and feature mapping is done in convolution layer. Pooling layer is mainly used to reduce number of parameters and computations. It selects the major features from feature map.

In flattening layer pooled feature map is converted into column. The reason to do this is we need to pass this as input to neural network and map can’t be given as input to neural network. And last layer of CNN is fully connected layer.

III. ALGORITHMS

Gaussian Blur Algorithm: Gaussian blur is the algorithm in which it is the result of blurring an image by a Gaussian function. It is also used as a pre-processing stage in computer vision algorithms in order to enhance image structure at different scales.

Steps of algorithm: Representing images as matrices of numbers.
Each number = intensity of a pixel.
More matrices = more colors.

Convex Hull (Jarvis's Algorithm or Wrapping): When a set of points is given, the smallest convex polygon which contains all the points is called as convex hull.

The simple idea of Jarvis's algorithm, start from the leftmost point (point with minimum X-coordinate) and then keep wrapping points in the anticlockwise (counterclockwise) direction.

For wrapping points orientation() method is used. If starting point is p, next point is selected as a point which beats all the other points in anticlockwise direction.

Edge Detection: Edge detection includes various methods which aims at identifying the points in digital image which changes brightness of image sharply, simply has discontinuities.

The points which changes brightness of image sharply are arranged (organized) onto a setoff curved line segments which are called as edges.

Edge detection is a fundamental tool in image processing, computer vision and machine vision, particularly in areas like feature detection and feature extraction.

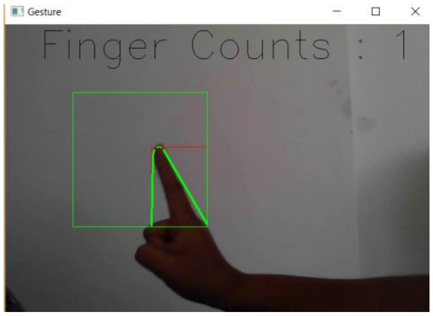
Otsu's method: Otsu's method is used for performing automatic image thresholding in computer vision and image processing. This algorithm returns the single intensity threshold that separate pixels in two different classes, foreground and background. The threshold is determined by maximizing inter-class intensity variance, or equivalently, by minimizing intra-class intensity variance.

Canny edge detection algorithm: In this algorithm edge detector is used as edge detector operator. To detect the wide range of edge in images multistage algorithm is used by edge detector operator .There are five steps in canny edge detection algorithm

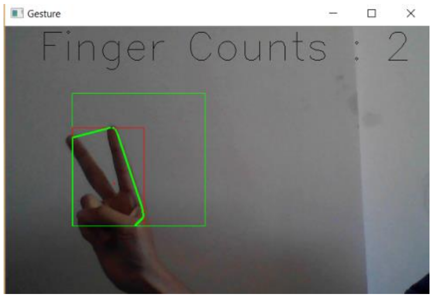
Initially apply Gaussian filter. To reduce noise and to smoothen the image Gaussian filter is used. Next intensity gradients of image are calculated. An edge in image may point to various directions. Hence four filters are used by canny algorithm to detect horizontal edges, vertical edges and diagonal edges. Values for first derivative in horizontal and vertical directions are returned by detection operator and form this edge gradient and direction is determined.

Now non-maximum suppression is applied. Non-maximum suppression is an edge thin technique. To find largest edge non-maximum suppression is applied. After application of non-maximum suppression double threshold is applied. After applying non-maximum suppression edge pixels give more accurate representation of edges but because of color variation and noise some edge pixel remain as it is and it is necessary to preserve the edge pixel with high gradient and filter out the pixels with low gradient value. And finally edges are finalized by suppressing weak edges and edges that are not connected to strong edges.

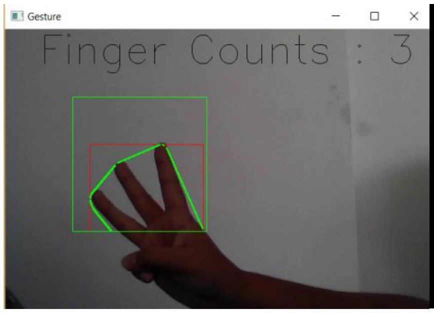
IV. RESULTS



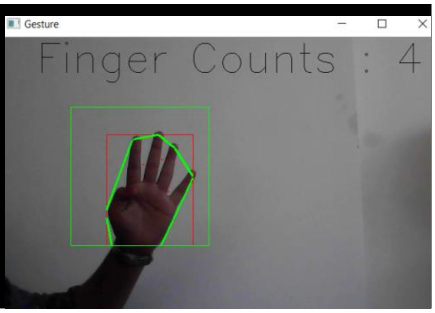
Finger Counts : 1



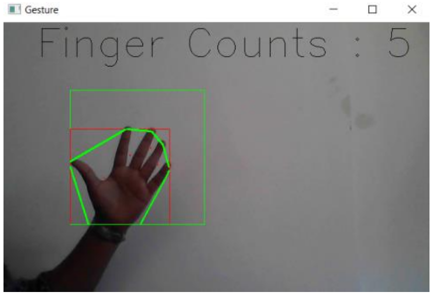
Finger Counts : 2



Finger Counts : 3



Finger Counts : 4



Finger Counts : 5

V. CONCLUSION

Through this paper we have proposed a novel implementation of sign language recognition on a desktop platform. Canny edge detection and otsu's techniques are used to segment the hand gesture from background captured by camera.

It must be noted that the test images used were captured under generally similar illumination and background conditions. The results may decrease somewhat if the background and illumination condition is changed. From our analysis of results, we found that sign language alphabets which are similar in appearance are prone to be misclassified and hence derive a lower accuracy. Whereas, visually distinctive sign language hand gestures tend to have high accuracy.

VI. FUTURE WORK

Future improvements to our framework may be to expand the application to wider content such as paragraph. Further more advanced algorithms may replace the current in place so as to improve the accuracy and increase the scope of the system.

VII. REFERENCES

1. M. M. Zaki, S. I. Shaheen, "Sign language recognition using a combination of new vision based features, " Pattern Recognition Letters, vol. 32, pp. 572–577, 2011.
2. C. Chuan, E. Regina, C. Guardino, "American Sign Language Recognition Using Leap Motion Sensor," in Proc. 13th International Conference on Machine Learning and Applications (ICMLA), Detroit, USA, 2014,pp. 541–544.
3. C. Dong, M. C. Leu, Z. Yin, "American Sign Language alphabet recognition using Microsoft Kinect," in Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, USA, 2015, pp. 44–52.
4. D. Aryanie, Y. Heryadi, "American sign language-based finger-spelling recognition using k-Nearest Neighbors classifier," in Proc. 3rd Inter-national Conference on Information and Communication Technology (ICoICT), Nusa Dua, Bali, 2015, pp. 533–536.
5. C. M. Jin, Z. Omar, M. H. Jaward, "A mobile application of American sign language translation via image processing algorithms," in Proc. IEEE Region 10 Symposium (TENSYMP), Bali, Indonesia, 2016, pp. 104–109
6. J. R. Pansare, M. Ingle, "Vision-based approach for American Sign Language recognition using Edge Orientation Histogram," in Proc. International Conference on Image, Vision and Computing (ICIVC),Portsmouth, UK, 2016, pp. 86–90.
7. V. N. T. Truong, C. Yang, Q. Tran, "A translator for American sign language to text and speech," in Proc. IEEE 5th Global Conference on Consumer Electronics, Kyoto, Japan, 2016, pp. 1–2.
8. M. M. Islam, S. Siddiqua, J. Afnan, "Real time Hand Gesture Recognition using different algorithms based on American Sign Language," in Proc. IEEE International Conference on Imaging, Vision Pattern Recognition (icIVPR), Dhaka, Bangladesh, 2017, pp. 1–6.
9. A. Joshi, H. Sierra, E. Arzuaga, "American sign language translation using edge detection and cross correlation," in Proc. IEEE Colombian Conference on Communications and Computing (COLCOM), Cartagena, Colombia, 2017, pp. 1–6.
10. M. Taskiran, S. Cimen, Z. G. Cam Taskiran, "The novel method for recognition of american sign language with ring projection and discrete wavelet transform, " World Journal of Engineering Research and Technology (WJERT), vol. 4, no. 1, pp. 92–101, 2018.