AMERICAN SIGN LANGUAGE GESTURE RECOGNITION

IMAGE PROCESSING
CSE4019
E1+TE1

PROJECT REPORT

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

NISHANT PANDEY- 17BCE0780
ASHWIN GOEL- 17BCE2073

Abstract

Inability to speak or listen is one of the biggest disabilities. For communication, such people prefer using sign language mostly. However, the biggest challenge with sign language is the inability to understand by normal people. Our aim is to tackle this problem by creating a vision based system which takes in a video of the person with disability and converts the sign language into English words/characters. By this, communication can be done in a more simpler and understandable way. We have achieved by taking a live feed of the user followed by hand segmentation and passing it into a trained CNN. The proposed technique is more accurate because of averaging of the result.

Keywords- *American Sign Language, CNN, Segmentation, Hand Gesture Recognition, Dilation, Erosion*

Introduction

Sign Gesture detection is one of the most trending topics and an active area of research. In the following system, a video stream is given is processed to detect human gestures. The two main processed done are skin recognition and contour based object detection.

Sign language can be used as a medium of communication for disabled people. And gesture detection can be used to process the human sign language into spoken/written language as needed.

In American Sign Language each alphabet A-Z, is assigned a unique hand gesture. Our goal is to design a Human Computer Interface that can understand human sign language simply by using vision based gesture detection and processing. In the project, a simple camera will be needed to capture the video stream, which will be further processed to get a minimal image needed for detection followed by a neural network which is trained on a large data-set for detecting such gestures. To make detection more accurate, the average result of 10 frames is considered.

Literature Review

[1] The paper deals with developing a system for automatic translation of static gestures of alphabets in American Sign Language. The system deals with images of bare hands, which is taken as an input. An image is processed and transformed to a feature vector that shall be compared with the feature vectors of a training set of signs. The system is implemented and tested using data sets of number of samples of hand images for each signs. Three feature extraction methods are tested and best one is suggested with results obtained from ANN. The system is able to detect American Sign Language signs with the accuracy of 92.33%.

[2] The purpose of this paper is to represent a real time HGR system based on American Sign Language (ASL) recognition with greater accuracy. This system acquires gesture images of ASL with black background from mobile video camera for feature extraction. In the processing phase, the system extracts features such as fingertip finder, eccentricity, pixel segmentation and rotation. For feature extraction, a new algorithm is proposed which basically combines K curvature and convex hull algorithms. It can be called “K convex hull” method which can detect fingertip with high accuracy. In the system, ANN is used with feed forward, back propagation algorithm for training a network using 30 feature vectors to recognize 37 signs of American alphabets and numbers properly which is helpful for HCI system. The system performs gesture recognition with an accuracy of 94.32%.

[3] The paper proposes a novel method to recognize symbols of the American Sign Language alphabet (A-Z) that have static

gestures. Many of the existing systems require the use of special data acquisition devices like data gloves which are expensive and difficult to handle. Some of the methods like finger tip detection do not recognize the alphabets which have closed fingers. A method is proposed where the boundary of the gesture image is approximated into a polygon with Douglas–Peucker algorithm. Each edge of the polygon is assigned the difference Freeman Chain Code Direction. Finger tips count along with difference chain code sequence as a feature vector is used. The matching is done by looking for either perfect match and in case there is no perfect match, sub string matching is done. The method efficiently recognizes the open and closed finger gestures.

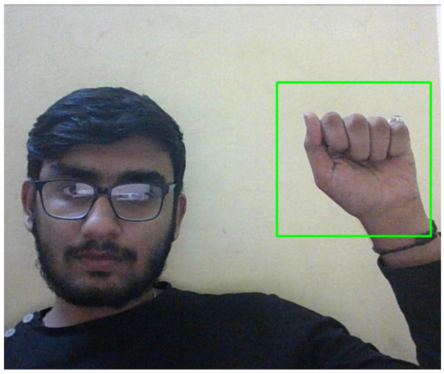
[4] In this work, real-time hand gesture system is proposed. This proposed method is divided into four stages such as image preprocessing, region extraction, feature extraction, feature matching. First stage converts captured RGB image into binary image using gray threshold method with noise removed using median filter and Guassian filter, followed by morphological operations. Second stage extracts hand region for getting region of interest and then "Sobel" edge detection is used. Third stage produces feature vector as centroid and area of edge, which will be compared with feature vectors of a training dataset of gestures using Euclidian distance in the fourth stage. This paper includes experiments for 26 static hand gestures related to A-Z alphabets. This gesture recognition system can achieve a 90.19% recognition rate in complex background.

Modules and Module Outputs

1. **Region Of Interest.**

   The region of interest is depicted by a green box , where the user needs to make a gesture to be detected.

   The region of interest is cropped out from the live frames.

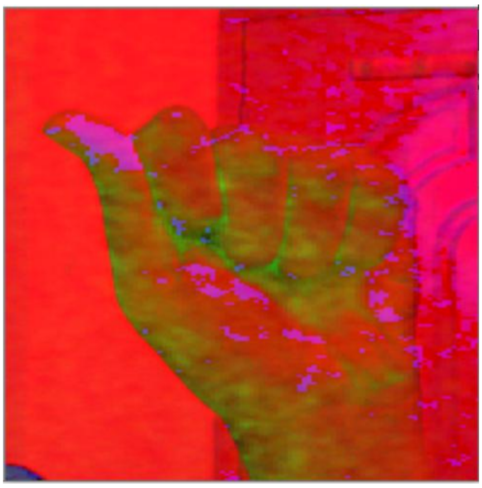


A frame from the live video

Region of interest separated out

2. **Colour Based Detection of hand**

The portion of hand is detected using color based detection by converting the image into HSV color scale. A mask is then applied to detect the skin color of the hand.



HSV converted Region of Interest

HSV mask

3. ENHANCEMENT OF THE MASK

The mask is smoothened with a 5X5 Gaussian Blur and thresholded to get a better segmented image of the hand.

Enhanced Image

4. CNN

The convolutional neural network contains the following layers-

A convolution layer which takes in a (128X128) image followed by a max-pooling layer with a pool size of (2X2).

There is another convolution layer with a max-pooling layer of pool size (2 X 2).

After this is the flattening layer .

The full connection has one Dense layer with 256 neurons with dropout followed by 26 output neurons.

The loss function used is categorical_cross-entropy and the optimizer is Stochastic gradient descent.

5. Training of the CNN

The CNN is trained on 45500 images of size (128X128) i.e 1750 images for each alphabet for 25 epochs with a batch size of 32 and validated on 6500 images.

Training image example for alphabet A

Algorithm

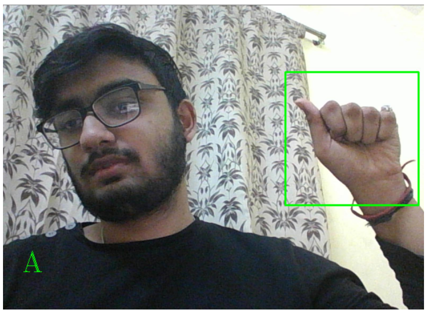
1. **Hand Detection**

    1. Taking live feed of the person
    2. Extracting the region of interest
    3. Converting the RGB image to HSV image
    4. Extracting the image of hand using color based object detection using skin color.
    5. Smoothening (applying Gaussian blur) to the image.
    6. Eroding and dilating the image of the extracted hand to remove noise.

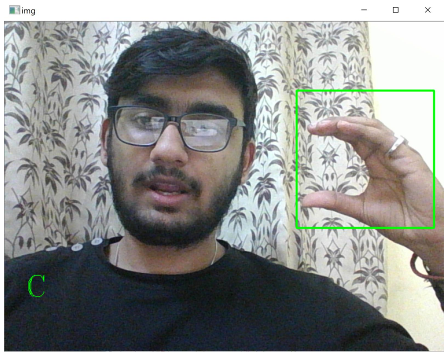
2. **Gesture Detection**

    1. Training a CNN from labeled images of hand
    2. Fine tuning the neural network
    3. Predicting the gesture(the American alphabet depicted from the gesture) of the extracted hand image through the neural network.

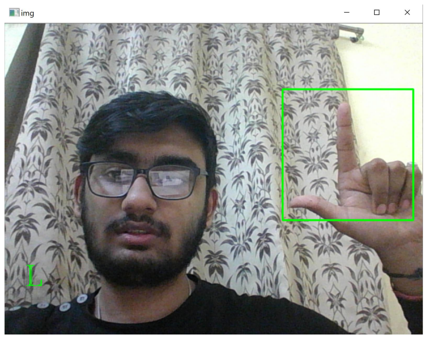
Experimental Results



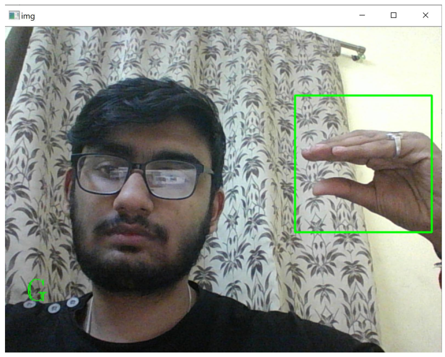
Sign representing letter: ***A***



Sign representing letter: ***C***



Sign representing letter: ***L***



Sign representing letter: ***G***

Dataset

The dataset contains labeled images of hand with about 2000 images each representing a single alphabetical sign.

Input Image

The input image is a (128X128) size segmented image of the region of interest from the live video capture.

Performance Evaluation

The system detects American Sign Language gestures with an accuracy of 98.06**%.**

To increase the accuracy of the result, averaging of the results of 10 consecutive frames is done. In this way the result doesn't fluctuate frequently.

Algorithm-metrics

Training accuracy = 99.03%

Training loss = 0.0277

Validation accuracy = 98.06%

Validation loss = 0.1309

	precision	recall	f1-score	support
A	0.05	0.05	0.05	250
B	0.02	0.02	0.02	250
C	0.02	0.02	0.02	250
D	0.04	0.04	0.04	250
E	0.02	0.02	0.02	250
F	0.04	0.04	0.04	250
G	0.03	0.04	0.04	250
H	0.04	0.03	0.04	250
I	0.06	0.06	0.06	250
J	0.03	0.03	0.03	250
K	0.04	0.04	0.04	250
L	0.02	0.02	0.02	250
M	0.03	0.04	0.04	250
N	0.04	0.04	0.04	250
O	0.03	0.02	0.02	250
P	0.05	0.05	0.05	250
Q	0.04	0.04	0.04	250
R	0.05	0.05	0.05	250
S	0.01	0.01	0.01	250
T	0.05	0.04	0.04	250
U	0.03	0.03	0.03	250
V	0.04	0.04	0.04	250
W	0.06	0.06	0.06	250
X	0.03	0.03	0.03	250
Y	0.05	0.05	0.05	250
Z	0.06	0.06	0.06	250
tal	0.04	0.04	0.04	6500

Conclusion

The paper proposes a system to detect the American Sign Language through gesture detection. This is achieved using a CNN trained from labeled data-set- labeled images of hand with respective signs. The system is able to recognize all symbols from A-Z using it's respective gestures in ASL. Using this method, results are proved to be more accurate because of averaging of the results of 10 frames. Limitations include using the system in complex background and also constant movement of hand due to which hand segmentation becomes difficult.

References

[1] Kulkarni, V. S., & Lokhande, S. D. (2010). Appearance based recognition of american sign language using gesture segmentation. *International Journal on Computer Science and Engineering*, *2*(03), 560-565.

[2] Islam, M. M., Siddiqua, S., & Afnan, J. (2017, February). Real time hand gesture recognition using different algorithms based on American sign language. In *2017 IEEE international conference on imaging, vision & pattern recognition (icIVPR)* (pp. 1-6). IEEE.

[3] Geetha, M., Menon, R., Jayan, S., James, R., & Janardhan, G. V. V. (2011, July). Gesture recognition for american sign language with polygon approximation. In *2011 IEEE International Conference on Technology for Education* (pp. 241-245). IEEE.

[4] Pansare, J. R., Gawande, S. H., & Ingle, M. (2012). Real-time static hand gesture recognition for American Sign Language (ASL) in complex background. *Journal of Signal and Information Processing*, *3*(03), 364.

[5] Liang, R. H., & Ouhyoung, M. (1998, April). A real-time continuous gesture recognition system for sign language. In *Proceedings third IEEE international conference on automatic face and gesture recognition* (pp. 558-567). IEEE.

[6] J. R. Pansare and M. Ingle, "Vision-based approach for American Sign Language recognition using Edge Orientation Histogram," *2016 International Conference on Image, Vision and Computing (ICIVC)*, Portsmouth, 2016, pp. 86-90.

[7] A. Thongtawee, O. Pinsanoh and Y. Kitjaidure, "A Novel Feature Extraction for American Sign Language Recognition Using Webcam," *2018 11th Biomedical Engineering International Conference (BMEiCON)*, Chiang Mai, 2018, pp. 1-5.