

ASL Fingerspelling Interpretaion

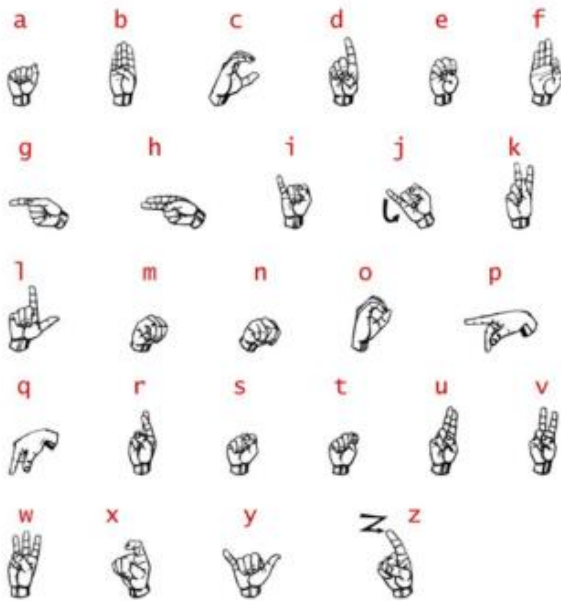
Sbera Catalin
Computer Science 30434

UTCN
catalin.sbera@yahoo.com

Abstract — This project will take images (live or not) of different hand gestures and try to identify what letter it corresponds to, displaying it along several other possible letters.

I. INTRODUCTION

The American Sign language (ASL) is one of the many forms of communication in the deaf communities. Compared to other forms, in which a whole word is represented by a gesture, this dialect will assign to each hand position 1 letter. Since the goal of the project is to deduct the letter from a static image, the letters J and Z will be omitted.



II. RELATED WORKS

A. Histogram of Centroid Distances (HOCD)

This method will take the image of a hand, generate the binary mask of the hand, extract an edge map by subtracting from the image the binary mask, calculate all the distances from the edge pixel from the centroid and compute a histogram with that data. Once that is done, the histogram will then be compared to a dataset of known histograms and try to find the “closest” one. Unfortunately, for some gestures, like for M and N, the gestures are too similar and this method will not provide the correct result all the time.

B. Gabor Filters

This filter is a similar one to the common gradient operator but is more flexible in terms of scale and orientation. Using a series of defined functions, their results are the extracted and considered as input in order to create a grid that will accurately calculate and classify the hand. Unfortunately, this algorithm struggles with determining the exact rotational differences on the segmented hand.

C. SURF

This method is based around extracting relevant data from the key point of each hand gesture. This data is then stored as words in a vector known as bag. Using this, histograms are being computed for n bins and then compared. This method is highly dependent on the illumination conditions, and thus, if for the training set the illumination is wildly different from the one from the test set, the result will not be accurate.

D. Hand Model

This method is based on the mathematical 20-degree-of-freedom model of the hand skeleton. The principle of the method is the following: for the image, the hand is segmented in order to get the silhouette, then, using an algorithm, a 3D model of the hand is modeled after it, which is then used to identify the corresponding letter. Unfortunately, finding an algorithm that accurately transposes an image from the 2D environment to an accurate 3D object is quite difficult.

E. Skin Segmentation

This is more of a preparation method that is done before applied other methods, like the Gabor filtering for example. It is, in essence, a filter, that based on known and computed RGB values of the human skin, will extract a binary mask of the hand.

III. CHOSEN METHOD

For this project I opted to only use the “HOCD” method since I will only translate a static image and this one is the most reliable in that context.

IV. IMPLEMENTATION FLOW

Firstly, we calculate and compute the binary mask of the hand, then we compute the centroid of that computed image. Next we extract an edge map of the hand with the help of the mask. Lastly, we compute the distances of each edge distance from the centroid and compute a histogram with those. Having now a database of histograms, when a new image is added, it will be compared to the database and choose which is the best approximation of it. There is a new option that allows the user to change the mode of the

program: From detecting the letter to telling the program what letter the image represents and saving it in the database.

Here is an example of the flow of the program. As one can observe, the edge is extracted then the centroid is defined. The program allows the user to toggle the display of those intermediate images.

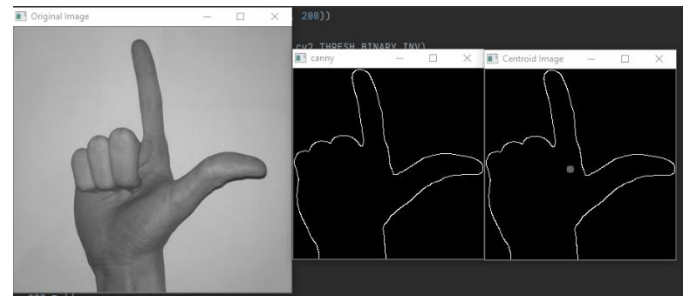
V. ENCOUNTERED PROBLEMS / RESULTS

Unfortunately, the edge detection algorithm is not optimal, thus the conditions for operating this program are quite strict: the hand must be at the center of the image, and the image must have a plain colored background in order for the edge to be correctly computed.

Other than this, there are a lack of samples on the internet for this topic, especially compared to the abundance of data for traffic sign detection, thus I had to add them manually.

VI. FUTURE IMPROVEMENTS / CONCLUSIONS

In the future, the images could firstly be going through a filter that will accentuate the color of the hands, especially the border of them, in order for the edge detection to work better.



Other than this an implementation of a algorithm that will exclude all the inner edges found in the inner part of the hands border plus one that can “approximate” the border of the hand if it has a few missing pixels would be a great improvement.

VII. REFENECES

[1] [ASL Fingerspelling Interpretation -Stanford University](#)

[2] [Project repository](#)