Hand Gesture Recognition for Sign Language

Syeda Miral Kazmi[1]

[1]Artificial Intelligence, Computing and Intelligent Design

[2]Computer Science Department

[3]Østfold University College
B R A Veien 4, 157 Halden, Norway

ABSTRACT

We have come to know a very genuine issue of sign language recognition, that problem being the issue of two-way communication i.e. between normal person and deaf/dumb. Current sign language recognition applications lack basic characteristics which are very necessary for the interaction with environment. Our project is focused on providing a portable and customizable solution for understanding sign language through an android app. The report summarizes the basic concepts and methods in creating this android application that uses gestures recognition to understand American sign language words. The project uses different image processing tools to separate the hand from the rest and then uses pattern recognition techniques for gesture recognition. A complete summary of the results obtained from the various tests performed is also provided to demonstrate the validity of the application.

Keywords: Image Processing · Sign Language · Recognizing Hand-gesture · Communication network· American Sign Language dataset · Machine learning.

INTRODUCTION

Human beings, on daily basis, communicate with each other and interact with computers using gestures. As a kind of gesture, sign language (SL) is the primary communication media for deaf people. In other words, sign language is a visual language that is used by deaf individuals as their tongue. It is achieved by combining shapes of hands, as well as, movement of it. Furthermore, it is also capable of recognizing arms or body, and facial expressions. Sign language is very important for deaf and dumb people for their social, emotional, and linguistic growth (Rogers, 2015). Two types of sign languages, Static and Dynamic, are suggested in this paper. Static sign only is able to recognize hand shape and orientation while dynamic signs recognize the direction of movement and sequence of gestures. Neural Networks have proved to work quite well for recognizing. For this purpose, we would use convolutional neural networks to have real-time translation. The current state of art technology such as YOLO net has proven to work quite well on detecting the hand gesture in real-time. Similar technology has also been employed in self-driving cars. This research could help to create a user interface. Which can be an android application or web interface that would capture hand gestures in real-time and then would be classified by a convolutional neural network. Sign language translation is one of the foremost difficult tasks within the field of Computer vision (Rogers, 2015). The thought is to tackle this challenge in real-time. Given a video feed, the idea is to sight the hand gesture in real-time then translate the corresponding sign into the corresponding linguistic communication. To achieve this task, we would be using American Sign language dataset. This research would propose a user interface of an android application that would capture hand gestures in real-time.

BACKGROUND

Problem Statement

Even though the deaf and dumb people can communicate with each other without having any issues, but there is a major challenge for them to communicate with normal people. Most normal people have not been educated about sign language. Therefore, they need a middle person to translate for both sides to make them understand each other. As communication is imperative, this issue makes an obstacle for the disabled people to compare with the normal. Therefore, there is need of efficient sign language translator to tackle those issues. This paper focuses on the research question "Is it possible to use image processing and machine learning to implement a hand gesture detection and translation to text in real-time".

Basic Operation

Firstly, liable, and calculated input for the gesture recognition system.

Image pre-processing: Secondly, separation and tracking is crucial to get needed information from input. Thus, it is necessary to be able to recognize the region of the foreground and split it from the background in each gesture image.

Feature vector extraction: Thirdly, the features (geometric and non-geometric) must be removed; these all features will be used at the time of testing operation (Mohandes, 2013).

Efficient classification algorithm: Classifier is used to identify the current presented testing gesture belongs to which trained class.

LITERATURE REVIEW

In order to have efficient communication between deaf and normal people, various work has been done in recent times by researchers. In this paper (Hogan, 2003), they argue that such a device should be feasible in the next few years, may provide immediate practical benefits for the deaf community. Finally, they describe initial recognition results of 94 percent accuracy on a 141 sign vocabulary signed in phrases of four signs using a one-handed glove-based system and hidden Markov models (HMMs). In this paper (Assaleh, 2005), they present a gesture recognition glove based on charge-transfer touch sensors for the translation of the American Sign Language. The prototype recognizes gestures for the numbers 0 to 9 and the 26 English alphabets, A to Z. The glove experimentally achieved, based on 1080 trials, overall detection accuracy of over 92 percent, which is comparable with current high-end counterparts. This paper (Mohandes, 2013) is to develop a Data Acquisition and Control system that translates the sign language into text that can be read by anyone. They developed a smart glove that captures the gesture of the hand and interprets these gestures into readable text. It is evident from the experimental results that gestures can be captured by a set of inexpensive sensors, which measure the positions and the orientation of the fingers. Recognition accuracy was 96 percent. Conference American Sign Language (Alzohairi, 2018) finger spelling translator based on skin segmentation and machine learning algorithms. The YCbCr color space is employed because it is typically used in video coding and provides effective use of chrominance information for modeling human skin color. Images depicting people of different ethnicity were used to simulate in order to measure the algorithm's performance. Then Convolutional Neural Network (CNN) is used to extract features from the images and Deep Learning Method is used to train a classifier to recognize Sign Language.

In terms of comparison between existing and proposed solution is that the existing system (Kakoty, 2018) can only detect the gesture if the background is black or white. In this proposed system hand gesture recognition works on every background with

average light. There was only one module in the existing system that is the gesture to text. In this proposed system, it's two-way communication because we are going to add both the module text to gesture and gesture to text.

METHODOLOGY

This section presents the method used for this study and how they were used. This paper focuses mainly on the iterative method for development. All the following developments and the direction of the paper is based upon the output from these review processes. We will use Agile methodology because it uses an incremental approach that can adapt to changes.

System Overview

In this paper (Elmahgiubi, 2015) using the calculation is created on top of the OpenCV library. All pictures caught are examined into the goal of 320x240px. in RGB design. The structure first examples RGB estimation of skin pixels of the underwriter for adjustment to actualize programmed identification of hand signal. Next, the distance variety between the hand motion and the camera is restricted to a specific reach. Use of a fitting scale-invariant procedure dispenses with the need to standardize the size of hand motion. We can use these conditions where the sure foreground and sure-foundation pixels can be resolved, also, consequently, the cultivated area developing technique is reasonable to be utilized in the division stage. Shrewd edge discovery is utilized to separate the hand's edges for highlighting extraction stage, they are used to restrict limit of cultivated district development to improve division exactness.

Hand Gesture to Normal Text

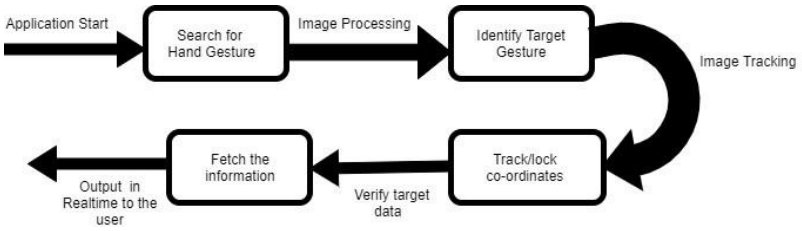

Figure 1. System workflow for translating of gesture to normal language

Calibration**.** For improvising strength of framework towards various clients and variety in bright conditions, an one-time alignment of skin pixels tone is done. A box

in red color is shown, and the endorser is provoked to catch a picture of the hand with the hand bigger than the red box. The greatest and least estimations of all R, G, and B channels are stored. In the web testing application, when the camera identifies skin pixels range in the focal point of the picture outline, it will naturally catch the outline. This strategy anyway is inclined to wrongly classify if foundation objects like skin tone are available (Jin, 2016).

Segmenting. The image in this stage is changed over from RGB shading space into a single channel grayscale picture. Canny edge location which is a method to distinguish and identify the presence of sharp discontinuities in a picture is applied to the picture to identify the edges. Next, a cultivated area developing strategy is utilized to isolate the hand district from the foundation. The locale developing technique is first utilized with an underlying seed point found on the definite foundations to deduct the foundation from the picture (Kane, 2012). The above cycle is iterated multiple times with three diverse seed focuses areas specifically on the upper left corner, upper right corner and base left corner, Next, the district developing strategy is applied with the seed point on the focal point of the picture. In the district developing cycle whereby the underlying seed point is put in the focal point of the picture, the past edge distinguished picture is utilized as a veil which serves to restrict the district developing limit. A locale developing condition is set to develop to the neighboring four pixels just if the pixels are non-zero, and henceforth it restricts the locale to just inside the limit of the endorser's hand. Ultimately, Canny edge identification is again applied to identify the edges in the sectioned hand picture (Rekha, 2011).

Feature Extraction. To recognize the hand gestures, Speeded Up Robust Features (SURF) will be used. As SURF is a patent local detector and descriptor, which helps to identify interest points or the important points of hand gestures. Additionally, the Hessian Matrix will be used to detect interest points. Apart from gesture recognition, various other tasks can be achieved by using SURF which ranges from object recognition, image registration, classification, or 3D construction. SURF, partly inspired by the SIFT descriptor, it is claimed standard version of SURF is several times faster and efficient than SIFT (Shahriar, 2018) since the application works by projecting an image of a hand which is showing a sign in real space, a good part of the project involves image processing Workflow of application.

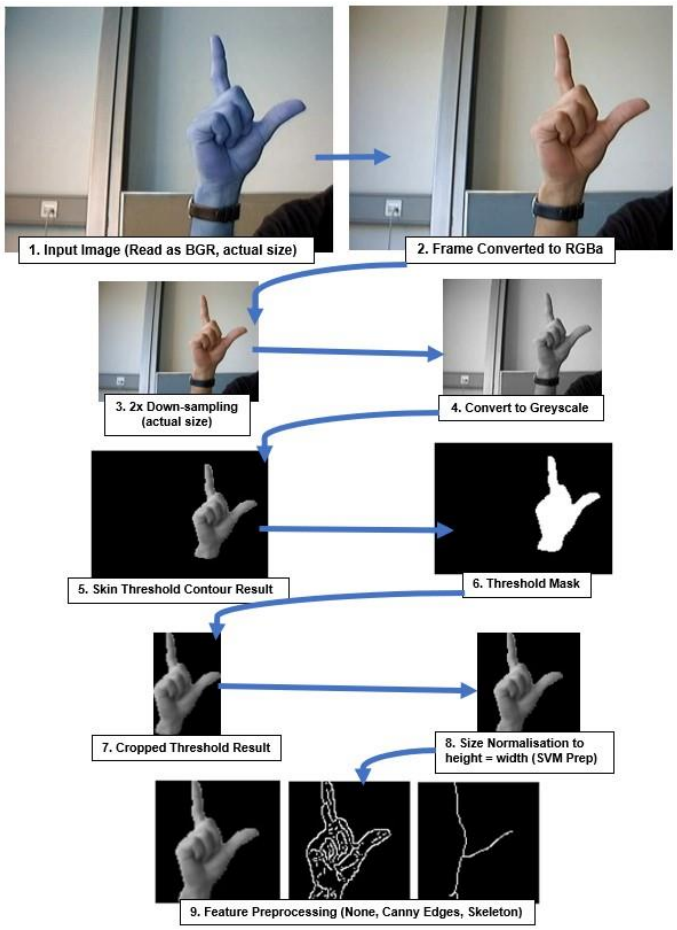

Figure 2. Processing and Initialization

Classification. To classify descriptors generated by SURF, SVM supervised machine learning technique will be considered. However, prior to the classification stage, to minimize the sum of squared Euclidean distances, the descriptors extracted by SURF are to be quantized in 16 clusters using the K-mean clustering technique. By doing this, each image will be represented by the frequency of occurrence of all clustered features. In our case of 16 clusters, this will give 16 classes of sign language gestures. Afterward, SVM will classify the sign of the languages, which could be translated into text or/and as speech output. (Jin, 2016).

Normal Text to Hand Gesture

Figure 3. System workflow for translating normal language to gesture

Data sets of gestures. The difficult task in machine learning when you want to train a good model for a prediction task is to find a large data set. We can use an already trained model as well (Kane, 2012). I proposed to use this dataset ww.lifeprint.com/asl101/pages-signs/. Through this dataset, we compare the input that was given by the user.

Working. The user can enter the alphabet, word, or sentence in the text area. The application will first try to recognize the word by processing it. Then it will match the text to the corresponding gesture. The next step is to retrieve the image through the given website linked and verify it. After verifying, it will display the gesture, if it does not get the word then it will spell the gesture of each alphabet of that word.

RESULTS

Testing and Evaluation:
This topic includes all the testing methodologies we have used to test our project. The most objective behind testing is to investigate whether the developed system meets the specified quality guidelines or not. System testing is significant for any system because it checks the capabilities of software and makes sure it is reliable.
Graphical user interface testing:
GUI testing is to figure out the defects present in an exceedingly software underneath test .This sort of test verifies that the application we tend to test will behaves as expected when a user performs a particular activity or enters a specific input in its activities and provide the proper program output .In our project the buttons in the android device were tested for actual response area size and by measuring the particular placement of the buttons on the Android device render.
Usability testing:
This type of testing technique which is utilized to check how simple and easy it is to use this system. This testing method is performed by the target users, who are going to be a part of the system. We make sure that our application camera runs in lower frame rates due to load on device to capture and detect and giving output at the same time. Almost 6 frames average

We have a small dataset to work with; 100 samples for each of 26 classes, after limited processing of images due to personal computer's capacity, we found the following relation:

Sample size	Validation Accuracy
25	28.5 %
50	35.8%
75	48.2%
100	62.5%

Figure 4:

Overfitting was being noticed during training our initial models on less data. Most probably it is because of small number of samples to train on leading to bad generalization and learning of the sample space (Brashear, 2003). This can be dealt by increasing the size of dataset to 100 samples, in our case the peak validation accuracy of 62.5 was achieved.

• The time complexity for training time = 20.73m.

• The time complexity for testing time = 2.53sec.

CONCLUSIONS

An application is developed for deaf and mute people through which they can communicate with a normal person who does not understand their gesture. This application can translate words and can translate gestures. We have developed an application which can detect hand gesture in Realtime. and display the recognized gesture into text on screen. The device's back camera is used to capture real time alphabetical gesture through OpenCV java camera enabled which gets the gesture in low frame rate by detecting a single hand gesture and testing that captured gesture with the classified datasets of gesture which is consist of 80:20 train and test dataset. Classification of gesture is processed fast therefore the framerate is low so it can detect and test the gesture smoothly. The Realtime gesture recognition works on every background with average light. In very high or very low light, it will be difficult to capture the gesture. The user can enter the letter, word or sentence in text area. The application will first try to recognize the word through the given website linked and will display the gesture, if it does not get the word then it will spell gesture of that word.

In the future we can enhance the feature of gesture to text. Right now, we are only converting gestures into alphabetic text but in future we can convert the words and sentences too. The other feature can also be enhanced by just not type the text to convert into gesture, but the user can speak the word, sentences so that the application shows the related gesture as an output.

REFERENCES

Alzohairi, R. A. (2018). Image based arabic sign language recognition system. International Journal of Advanced Computer Science and Applications (IJACSA),.

Assaleh, K. &.-R. (2005). Recognition of Arabic sign language alphabet using polynomial classifiers. EURASIP Journal on Advances in Signal Processing, 1-10.

Brashear, H. S. (2003). Using multiple sensors for mobile sign language recognition. Georgia Institute of Technology.

Elmahgiubi, M. E. (2015). Sign language translator and gesture recognition. 2015 Global Summit on Computer & Information Technology (GSCIT) IEEE, 1-6.

Hogan, K. (2003). Can't get through: eight barriers to communication. Pelican Publishing.

Jin, C. M. (2016). A mobile application of American sign language translation via image processing algorithms. 2016 IEEE Region 10 Symposium (TENSYMP), 104-109.

Kakoty, N. M. (2018). Recognition of sign language alphabets and numbers based on hand kinematics using a data glove. Procedia Computer Science,.

Kane, L. &. (2012). Towards establishing a mute communication: An Indian Sign Language perspective. 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI), 1-6.

Mohandes, M. A. (2013). Recognition of two-handed arabic signs using the cyberglove. Arabian Journal for Science and Engineering,, 669-677.

Rekha, J. B. (2011). Shape, texture and local movement hand gesture features for indian sign language recognition. 3rd International Conference on Trendz in Information Sciences & Computing (TISC2011), 30-35.

Rogers, Y. J. (2015). Interaction Design: Beyond Human - Computer Interaction. Wiley, 4th ed. John Wiley Sons Inc.

Shahriar, S. S. (2018). Real-time american sign language recognition using skin segmentation and image category classification with convolutional neural network and deep learning. ENCON 2018-2018 IEEE Region 10 Conference, 1168-1171.