American Sign Language Translator

W.C.C. Premarathna[1], T. Arudchelvam[2]

[1, 2]Department of Computing and Information Systems, Wayamba University of Sri Lanka, Sri Lanka
[#] W.C.C. Premarathna; premarathna142083@wyb.ac.lk

Abstract— *The Sign language is very important for people who have hearing and speaking deficiency generally called Deaf or Muted people. Every normal human being sees, listens, and reacts to surrounding. But those are unlucky individuals who does not have this important blessing. Such individuals, mainly deaf and dumb, they depend on communication via sign language to interact with others. However, communication with ordinary individuals is a major impairment for them since not every typical people comprehend their sign language. This paper proposes an application which would help in recognizing the different signs which is called ASL (American Sign Language) by using Python, OpenCV, Tensorflow and Keras. The images are of the palm side of hand and are loaded at runtime. The method has been developed with respect to single user at a time. The real time images called as training data is captured first and then stored in directory. Then feature extraction will take place to identify which sign has been articulated by the user. Finally a CNN (Convolution Neural Network) model which is used sequential classifier and RELU (Recurrent Linear Units) activation function was created and saved as a json file. The comparisons will be performed and then after comparison the result will be produced in accordance through matched key points from the input image to the image stored for a specific letter already in the json model. There are 41 signs in ASL corresponding to each 26 English alphabet, 0 – 9 numbers and some simple words also. This model provided with 95% accurate results for input images captured at many possible angle and distance in a pleasant environment.*

Keywords— **American Sign Language (ASL), Python, OpenCV, Tensorflow, Keras, Rectified linear unit (RELU), Convolution Neural Network (CNN), British Sign Language (BSL), Region of Interest (ROI).**

I. INTRODUCTION

Each individual utilize to a language to communicate with others. Sign language is basically utilized by hearing impaired people to communicate with each other, developed by deaf communities. Communication through signing language is well organized gestural language using both non-manual and manual correspondence. Non manual signals are outward appearance, body orientation, head movement, stance and eye blinking. Manual signals are incorporated with movement and orientation of fingers or hand that passes on typical significance [Ahmed, 2015].

Communication with typical individuals is a major impediment for them since not every ordinary people comprehend their gesture based communication. To overcome this problem, sign language recognition system is expected to assist the deaf or muted people to communicate with normal people. "One of every five people who are deaf in the world"[WHO, 2018], making it the country with the largest number of Deaf, and perhaps also the largest number of sign language users. ASL is considered as the predominant sign language of Deaf communities in the United States. ASL has propagated widely via schools for the deaf and deaf community organizations. Nowadays, it has been spread in many countries. Hence, the design of the sign language recognition system will be much relevant and based on the ASL, in order to suit the local people environment. [Mitchell, 2004]

A. What is ASL?

American Sign Language (ASL) is a well-organized and noncomplex language that signs are made by moving of the fingers or hands. It is the main language of many North Americans who are completely deaf or muted communities. Nowadays ASL is spread all over the world and many deaf schools and deaf communities uses ASL as a primary language.

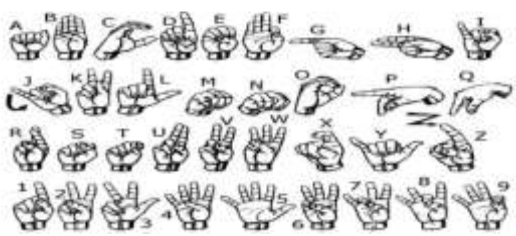


Figure 1. Letters of the alphabet and numbers in American Sign Language.

But ASL is not a universal sign language. Different countries or regions have different sign languages. For example, BSL (British Sign Language) is a different language from ASL. But ASL has many signs that are quiet similar to other languages.

II. LITERATURE REVIEW

A. Data-glove approach.

The data-glove approach is an approach that utilizes a unique assembled electronic glove, which has in-fabricated sensors that is utilized to distinguish the hand gestures or postures. Most of commercial sign language translation systems use the data-glove approach, because it is very easy to acquire data on the bending of finger and 3D orientation of the hand using gloves [F. Yin, 2016]. The data glove consists of ten flex sensors, two on each finger. [Hoffman, 2016]

Figure 2: Data glove with flex sensors

B. Visual-based approaches

With late progression in PC and data innovation, there has been an expanded regard for visual-based methodology. Images of the signer is captured by a camera and video processing is done to perform acknowledgment of the sign language. For the glove which is crafted for specific tasks, the signer is furnished with color-coded gloves [T.T. Swee, 2007]. The colour will give the extraction of information from the images of the signer through colour segmentation. These gloves are essentially normal pair of gloves with particular shading on every fingertip and palm. Intergraded webcam is used to collect images from the signer in type of still images and video streams in RGB (red-green-blue) shading [Y. Madhuri, 2013].

These visual-based approaches are fundamentally minimizing the equipment necessities and cost. However, this kind of systems are just suitable and viable for recognizing alphabets and numbers, as opposed to perceiving sign gestures. [T.T. Swee, 2007]

Figure 3: Color-coded gloves

C. Virtual-based Button Approach

The main aim of virtual button approach is to create a function of a virtual button which generates button events, a press and discharge, by perceiving hand motions of holding and discharging individually. This virtual button is also used to identify different sorts of gesture and generate proper command which is suitable for postures. The virtual button approach uses patterns of the wrist shape. It can perceive a sequential movements of hand. By using small sized IR optic sensors, the patterns of hand gestures or postures can be recognized by moving fingers. These patterns are cooperated finger or hand movement. The IR emitter and IR optic sensor are mounted on the bottom of the wrist because the area comprises of finger flexor tendons that directly respond to the finger movements. These sensors produce voltage values according to the amount of IR radiation. In the system, this sensor is used to monitor different patterns of the wrist shape resulting from the movement of finger flexor tendons in the wrist when fingers are moving [Ahmed, 2015].

III. METHODOLOGY

This project consist of four major parts which are namely Image Acquisition (Collect Data), Data training, predict the Output and Gesture Recognition (a support to the model) which are shown in Figure 4.

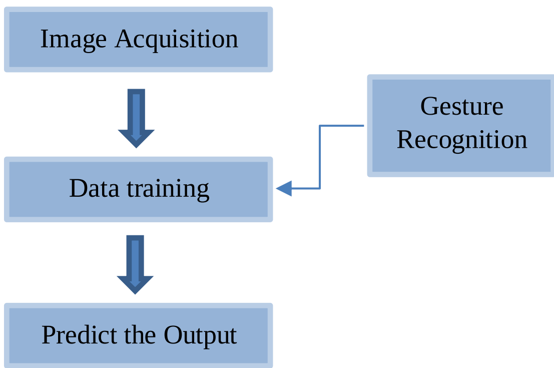


Figure 4. Major parts of ASL Translator.

A. *Image Acquisition.*

The first step of the project is Image Acquisition i.e. collect the training data set. Train data set is used to implement the CNN and test data set is used to check the accuracy of the CNN model.

The images will be captured through basic code of opening a webcam through OpenCV and then capturing the image through frames per press which will be stored for a specific letter in the directory. For an example, an image for sign for letter "A" will be captured and stored in the directory when the user press the letter "A" key in the key board. If user press the Esc key, image acquisition process will be terminated. Following figure 5 and Figure 6 show the interfaces of the image acquisition process for both training and testing data sets.

Figure 5. User interface of image acquisition for training data set.

Figure 6. User interface of image acquisition for testing data set.

B. *Data Training.*

The Data Training part is again divided into to three major steps.

- Step 1 - Building the CNN.

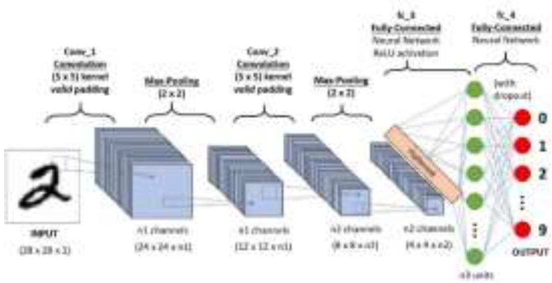


Figure 7. A CNN sequence to classify a sign.

- Step 2 - Preparing the training /testing data and training the model.
- Step 3 - Saving the model.

C. *Predict the Output.*

A quite similar interface which is used to collect training and testing data is used to acquire the input sign language. Here also have a region called as ROI. User put the hand in this section and provide some postures by hand. The web camera captures the hand by using a code segment which is coded by OpenCV at a frame rate of 24. Each image is continuously compared with the model and gives a predicted output as a text in the top of left corner in the interface.

Figure 8. Predicted output for number 1

Figure 9. Predicted output for letter "A"

Figure 10. Predicted output for words

D. *Gesture recognition.*

In this method, some mathematical approaches are used to find the gestures. Major steps are given below:

i. Define the region of interest
ii. Define range of skin color in HSV
iii. Extract skin color from the image
iv. Make convex hull around the hand.
v. Define area of convex hull and area of hand;

$area_{hull}$ = area of the convex hull
$area_{hand}$ = area of the hand

vi. Calculate the percentage of area not covered by hand in convex hull

$$area_{ratio} = \left(\frac{area_{hull} - area_{hand}}{area_{hand}}\right) * 100$$

vii. Find the number of triangles which can be made from the fingers of hand.

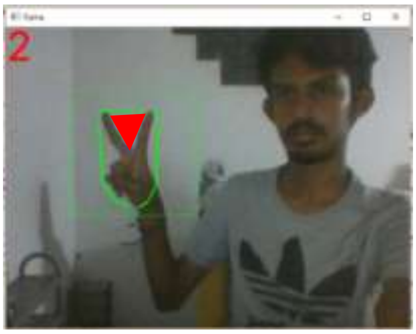


Figure 11. Number of triangles which can be made from fingers

viii. Find length of all sides of triangle.

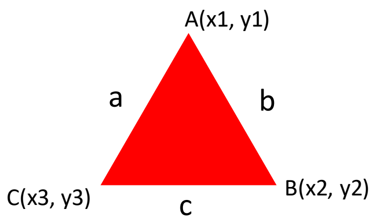


Figure 12. Coordinates of triangle

$$a = \sqrt{(x2 - x1)2 + (y2 - y1)2}$$

$$b = \sqrt{(x2 - x3)2 + (y2 - y3)2}$$

$$c = \sqrt{(x1 - x3)2 + (y1 - y3)2}$$

ix. Apply cosine rule here to find the angle

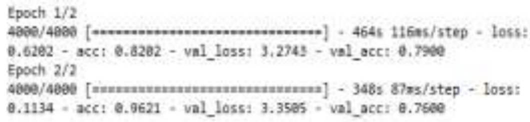
$$angle = \cos^{-1}\left(\frac{b2 + c2 - a2}{2 * b * c}\right) * 57$$

x. Ignore angle > 90 and ignore points very close to convex hull (angle >30) because they generally come due to noise.

IV. RESULTS

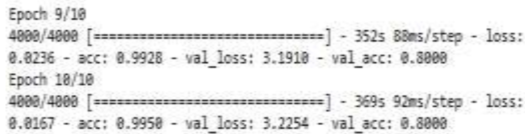
A. Gesture recognition.

Since 4000 + 200 images used 10 as the epoch number in order to train the model. It gave a higher level of value accuracy with 0.9950 rather than the lower number of epoch number. Figure 13 and Figure 14 show the accuracy number with epoch number as 2 and 10.



```
Epoch 1/2
4000/4000 [==============================] - 464s 116ms/step - loss:
0.6202 - acc: 0.8202 - val_loss: 3.2743 - val_acc: 0.7900
Epoch 2/2
4000/4000 [==============================] - 348s 87ms/step - loss:
0.1134 - acc: 0.9621 - val_loss: 3.3505 - val_acc: 0.7600
```

Figure 13. Maximum value accuracy with epoch number 2 (0.9621)



```
Epoch 9/10
4000/4000 [==============================] - 352s 88ms/step - loss:
0.0236 - acc: 0.9928 - val_loss: 3.1910 - val_acc: 0.8000
Epoch 10/10
4000/4000 [==============================] - 369s 92ms/step - loss:
0.0167 - acc: 0.9950 - val_loss: 3.2254 - val_acc: 0.8000
```

Figure 14. Maximum value accuracy with epoch number 10 (0.9950)

B. Prediction Accuracy.

When comparing input images with CNN model, the model will construct a prediction table according to the resultant values and sort those value in descending order as the highest resultant value to be in the top. Following Figure 15 clarify the prediction values in a histogram approach with highest probabilities are in left most side. Values will be decreased from left to right.

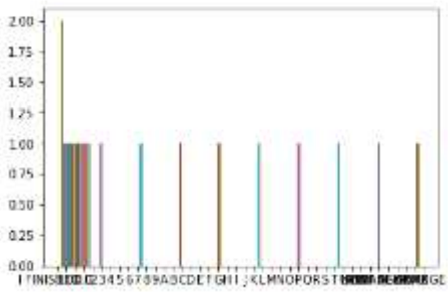


Figure 15. Histogram approach with highest probabilities.

C. The support to the model.

Although model was trained with over 4000 images, sometimes, it gives incorrect predictions. And also some features of the image may be lost due to the grey scale conversion and threshold process. Therefore, gesture recognition which was described in section III.D, gives considerably adequate support to the model.

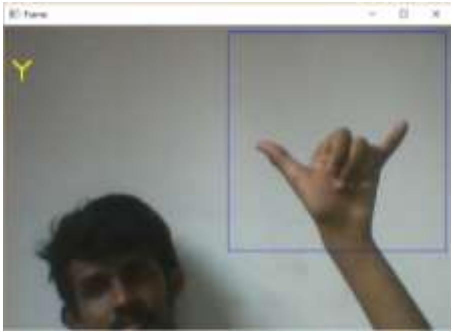


Figure 16. With the support of gesture recognition (correct).

Figure 17. Without the support of gesture recognition (incorrect).

In Figure 16 model predict the correct output with the support of gesture recognition. In Figure 17 for the same gesture model predict the incorrect output.

V. DISCUSSION AND CONCLUSION

This paper is about a system that can support the communication between deaf or muted and ordinary people. The aim of the work is to provide an output as a text without knowing sign language. The program has four major parts. Firstly, the image acquisition part is used to collect, training and testing data. It takes images of hand gestures, converts it into grey scale and stores in a specific directory. Secondly, the training part uses CNN building methods to create a model. The accuracy of the model was increased with the increment of the epoch number (Figure 13 and 14). Thirdly, output prediction part uses the created model to give possible output as a text to input sign language. Finally, gesture recognition part is used in order to give a support to the model.

To predict correct outputs, the model needs a clear and pleasant (without sunlight) environment. Hence in some environmental conditions, model gave incorrect results. Model was trained with respect to one person. Therefore, the application cannot operate with several users. Sometimes CNN model gives some conflicts, when user shows the signs in left hand. The reason for this problem is data set for training process is collected only for right hand.

VI. ACKNOWLEDGEMENT

I hereby would like to take this opportunity to thank all persons who has involved generously in helping me and assisting me while I was completing the final year research project. I would firstly to express my deepest gratitude and thanks to my project supervisor, Mr. T. Arudchelvam for his undivided support morally and physically, assistance, guidance, tolerance, which proved to be invaluable as to completion my final year research project. We would also like to thank our department for giving us the freedom to pursue this project. A special thanks to my family. I am very thankful to my mother and father for all of the sacrifices that you've made on my behalf. Also thanks to my friends and everyone, that have been contributed by supporting my work and help myself during the research project progress till it is fully completed.

VII. REFERENCES

Ahmed Zaki Shukor, Muhammad Fashmi Miskon, Muhammad Hemman Jamaluddin, Fariz Bin , "A new data glove approach for malaysian sign language detection", 2015 IEEE International Symposium and Intelligent Sensors (IRIS 2015).

C. Preetham, G. Ramakrishnan & S. Kumar, "Hand Talk-Implementation of a Gesture Recognizing Glove", 2013 Texas Instruments India Educators' Conference (TIIEC), Bangalore, April 2013, pp 328-331.

Mitchell RE, Karchmer MA., "Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States Sign Language Studies." 2004; 4(2):138-163.

R. Akmeliawati, "Real-Time Malaysian Sign Language Translation using Colour Segmentation and Neural Network", Proc. of the IEEE International Conference on Instrumentation and Measurement Technology 2007, Warsaw, 2007, pp 1-6.

T.T. Swee, A.K. Ariff, S.-H Salleh, S.K. Seng & L.S. Huat, "Wireless Data Gloves Malay Sign Language Recognition System", 2007 6th International Conference on Information, Communications & Signal Processing, 2007, pp 1-4.

WHO – World Health Organization, Deafness and hearing loss,https://www.who.int/newsroom/factsheets/detail/deafnes s-and-hearing-loss.

Y. Madhuri, G. Anitha & M. Anburajan " Vision-based Sign Language Translation Device", 2013 International Conference on Information Communication and Embedded Systems (ICICES), 2013, pp 565-568.

Hoffman HJ, Dobie RA, Losonczy KG, Themann CL, Flamme GA., "Declining Prevalence of Hearing Loss in US Adults Aged 20 to 69

Years”, JAMA Otolaryngology – Head & Neck Surgery. December 2016.

F. Yin, X. Chai, and X. Chen. Iterative Reference DrivenMetric Learning for Signer Independent Isolated Sign Language Recognition. In European Conference on Computer Vision (ECCV), 2016.