A Project Report on

Gesture-based real-time recognition system for Indian Sign Language (ISL)

Submitted by

Taj Shah (4119006)
Hamza Ansari (4119007)
Advait Chavan (4119008)
Aakash Gupta (4119009)

in partial fulfillment for the award of the degree

BACHELOR OF ENGINEERING

in

Electronics Engineering

Under the Guidance of

Prof. Javed Taili

M. H. Saboo Siddik College of Engineering

University of Mumbai

2022 - 2023

CERTIFICATE

This is to certify that Students Taj Shah, Hamza Ansari, Advait Chavan, Aakash Gupta are the bonafide students of M. H. Saboo Siddik College of Engineering, Mumbai. They have successfully carried out the project titled **"Gesture-based real-time recognition system for Indian Sign Language (ISL)"** in partial fulfilment of the requirement of B.E. Degree in Electronics Engineering of Mumbai University during the academic year 2022-2023. The work has not been presented elsewhere for the award of any other degree or diploma priorto this.

(Prof. Javed Taili)

(Prof. Amar Pawade)

(Prof. Jilani Sayyad)
ELEX HOD

(Dr. Ganesh Kame)
Principal

Project Report Approval for B.E.

This project entitled ***'*Gesture-based real-time recognition system for Indian Sign Language (ISL)*'*** by **Taj Shah, Hamza Ansari, Advait Chavan, Aakash Gupta** is approved for the degree of Bachelor of Engineering in Electronics from University of Mumbai.

Examiners

1. ------------------

2. ------------------

Date:

Place:

ACKNOWLEDGEMENT

We are thankful to a number of individuals who have contributed towards our final year project and without their help; it would not have been possible. Firstly, we offer our sincere thanks to our project guide, Javed Taili for his constant and timely help and guidance throughout our preparation.

We are grateful to Er. Amar Pawade project coordinator ELEX department, for his valuable inputs to our project. We are also grateful to the college authorities and the entire faculty of the ELEX department for their support in providing us with the facilities required throughout this semester.

We are also highly grateful to Prof. Jilani Sayyad of Department (ELEX), Principal, Dr. Ganesh Kame and Director Dr. Mohiuddin Ahmed for providing the facilities, a conducive environment and encouragement.

Signatures of all the students in the group

(Taj Shah)

(Hamza Ansari)

(Advait Chavan)

(Aakash Gupta)

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included; we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in this submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Signatures of all the students in the group

(Taj Shah)

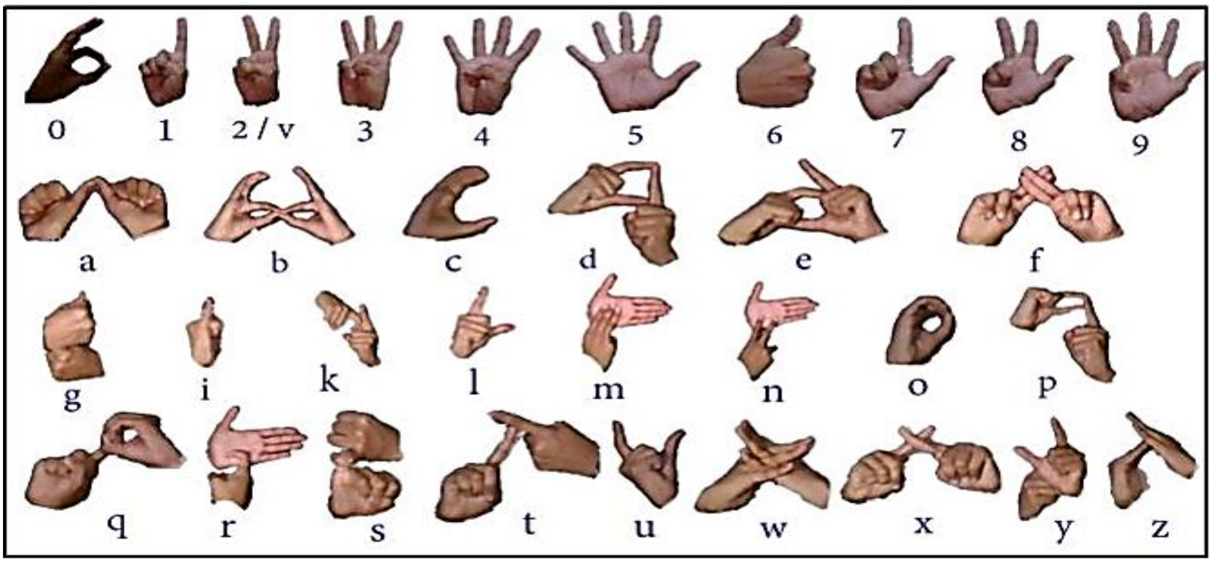
(Hamza Ansari)

(Advait Chavan)

(Aakash Gupta)

Abstract

This report presents a system that can recognize Indian Sign Language (ISL) hand poses and gestures performed by deaf and dumb people in real time and interpret/translate them into them English text readable to normal people for better and efficient communication within the society. The existing solutions are either of high cost of purchase and maintenance which cannot be afforded by most of the deaf and dumb people of India in particular. Also, the existing solutions have a low accuracy and low speed of operation which will have an effect on the real-time analysis, interpretation and translation of the sign language to a language which shall be understandable by the physically normal people. Our proposed system has the ability to overcome both the above-mentioned problems i.e. Accurate Results and Time Factor. It can identify 35 ISL gestures along with some hand poses. Sign Language is captured from a smartphone camera and its frames are transmitted to a remote server for processing. The use of any external hardware (such as gloves or the Microsoft Kinect sensor) is avoided, making it user-friendly



Contents

Chapter 1

Introduction

1.1 Motivation

Indian Sign Language (ISL) is a sign language used by hearing and speech impaired people to communicate with other people. The research presented in this report pertains to ISL as defined in the Talking Hands website [1]. ISL uses gestures for representing complex words and sentences. It contains 33 hand poses including 10 digits, and 23 letters. Amongst the letters in ISL, the letters ‘h’, ‘j’ are represented by gestures and the letter ‘v’ is similar to digit 2. The system is trained with the hand poses in ISL as shown in Fig. 1. Most people find it difficult to comprehend ISL gestures. This has created a communication gap between people who understand ISL and those who do not. One cannot always find an interpreter to translate these gestures when needed. To facilitate this communication, a potential solution was implemented which would translate hand poses and gestures from ISL in real-time. It comprises of an Android smartphone Camera/ Desktop or PC Camera to capture hand poses and gestures, and a server to process the frames received from the smartphone camera/PC camera. The purpose of the system is to implement a fast and accurate recognition technique.

1.2 Problem Statement

Understanding the exact context of symbolic expressions of deaf and dumb people is the challenging job in real life until and unless it is properly specified.

To fill the bridge between the communication process between the deaf and dumb people and normal people using the most effective, fast and low-cost method.

Chapter 2

Literature Review

There has been considerable work in the field of Sign Language Recognizer with Novel approaches towards Gesture Recognition.

Different methods such as use hardware components of Gloves or Microsoft Kinect sensor for tracking hand, etc.

2.1 Literature Review on Sign Language Recognizer (SLR) using hardware

A Microsoft Kinect sensor is used in [2] for recognizing sign languages. The sensor creates depth frames; a gesture is viewed as a sequence of these depth frames. But, the sensor of Microsoft Kinect used is almost for 150$ (INR 4342) which is not affordable by most of the dumb and deaf people in India.

T.Pryor et al [3] designed a pair of gloves, called SignAloud which uses embedded sensors in gloves to track the position and movement of hands, thus converting gestures to speech, But, it is not feasible and affordable for the dumb and deaf people in India.

R. Hait-Campbell et al [4] developed MotionSavvy, a technology that uses Windows tablet and Leap Motion accelerator AXLR8R to recognize the hand, arm skeleton.

Sceptre [5] uses Myo gesture-control armbands that provide accelerometer, gyroscope and electromyography (EMG) data for signs & gestures classification.

These hardware solutions provide good accuracy but are usually expensive and are not portable. Our system eliminates the need of external sensors by relying on an Android phone camera/ PC camera

2.2 Literature Review on Sign Language Recognizer (SLR) using software

For software-based solutions, there are colored glove based and skin color-based solutions.

R. Y. Wang et al [6] have used multi-colored glove for accurate hand pose reconstruction but the sign demonstrator, while demonstrating the sign language, has to wear this each time.

Skin color-based solutions may use RGB color space with some motion cues [8] or HSV [9, 10, 11], YCrCb [12] color space for luminosity invariance.

G. Awad et al [13] have used the initial frames of the video sequence to train the SVM for skin color variations for the further frames. But to speed up the skin segmentation, they have used Kalman filter for further prediction of position of skin colored objects thus reducing the search space.

Z. H. Al-Tairi et al [14] have used YUV and RGB colour space for skin segmentation and the color ranges that they have used handles good variation of
people's races.

Chapter 3

Proposed Design Methodology

Using an Android smartphone or a desktop, gestures and signs performed by the person using ISL are captured and their frames are transmitted to the server for processing. To make the frames ready for recognition of gestures and hand poses, they need to be pre-processed. The pre-processing first involves converting the original image to its gray scale format.

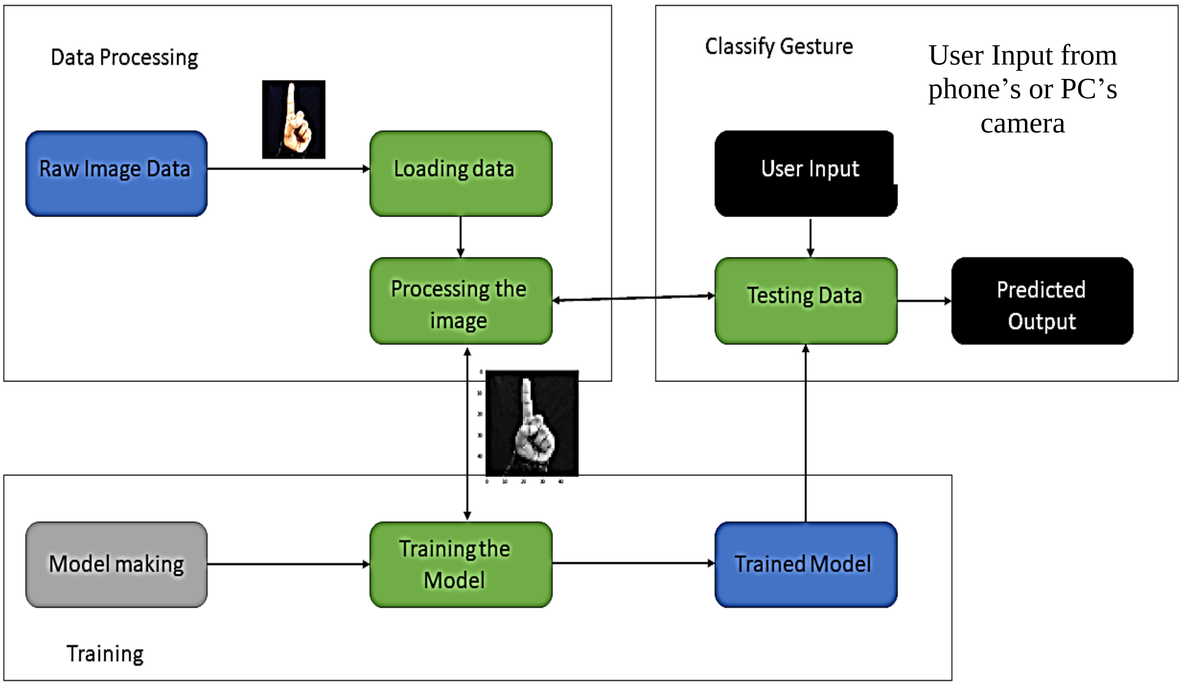


Fig. 2. Flow diagram for Gesture Recognition.

The image converted to grayscale is then processed using the Inception V3 Convolutional Neural Network Deep Learning Model which is used for image analysis where the image are analyzed on the basis of their weights. We have used '**Mixed 7**' as an output layer for inception V3. Then by using Softmax and Relu functions of Deep Learning the layers are being trained. For compiling we have used SGD optimizer of inception V3 Deep Learning Model. Stochastic gradient descent (SGD) is the simplest update: the weights are nudged in the negative gradient direction. Despite its simplicity, good results can still be obtained on some models. The updates dynamics can be written as:

$$w_{k+1} = w_k - \alpha \nabla f(w_k)$$

An overview of this process is described in Fig. 2.

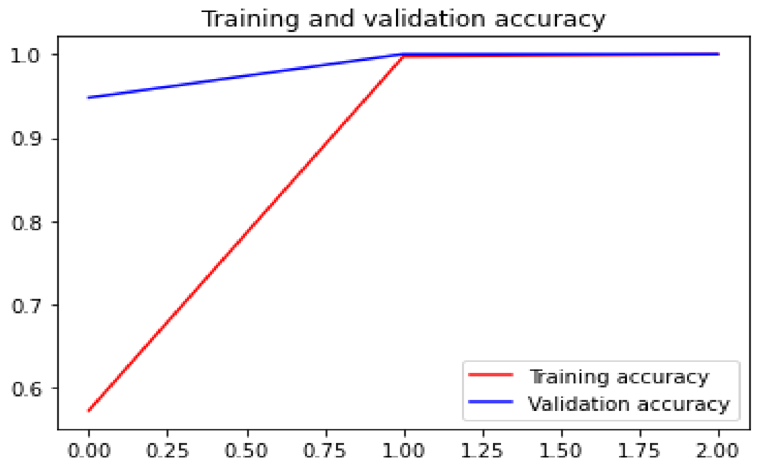
3.1 Dataset.

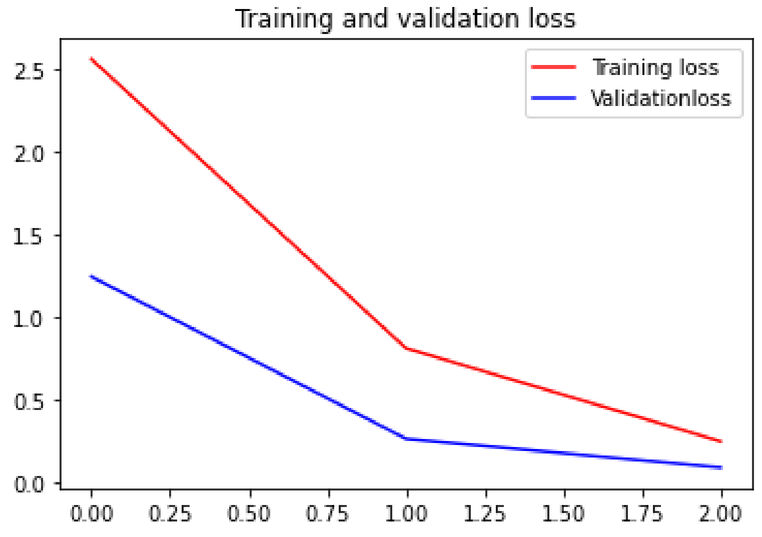
For the digits 0 to 9 in ISL, an average of 1200 images per digit were captured. For letters in ISL, about 1200 images per letter were captured. The dataset contains a total of 42,000 images. All the images consist of the sign language using one hand(for digits) and 2 hands(for letters). Most of these images were captured from an ordinary webcam and a few of them were captured from a smartphone camera. The images are of varying resolutions.

Chapter 4

Simulation and Experimental Results

The Graphs for Training and Validation Accuracy along with Training and Validation Loss were plotted for the model which is being trained using InceptionV3 Convolutional Neural Network model which is used for Image Analysis and Object Detection.





Chapter 5

Conclusion

5.1 Conclusion

In the report, we have trained the model using Inception V3 Deep Learning Model and from Graphs we can conclude that we got good results. Also we got a Accuracy of 97.89% for the training v/s validation accuracy graphs and training v/s validation loss graph.

5.2 Future Scope-

Future work will focus on taking into consideration entire upper body part of the human body (section of body above the waist) and then filtering out the facial part from the frame of the image and only considering the hand movement for predicting and translating to English language in real time.

References

[1] TalkingHands.co.in, “Talking Hands,” 2014. [Online]. Available: http://www.talkinghands.co.in/. [Accessed: 20-Sept-2022].

[2] A. Agarwal and M. K. Thakur, “Sign Language Recognition using Microsoft Kinect,” Sixth International Conference on Contemporary Computing (IC3), September 2013.

[3] MailOnline, ‘'SignAloud gloves translate sign language gestures into spoken English,” 2016. [Online]. Available: http://www.dailymail.co.uk/sciencetech/article-3557362/SignAloud gloves-translate-sign-language-movements-spoken-English.html. . [Accessed: 25- Sept- 2022].

[4] Alexia. Tsotsis, “MotionSavvy Is A Tablet App That Understands Sign Language,” 2014. [Online]. Available: https://techcrunch.com/2014/06/06/motionsavvy-is-a-tablet-app-that understands-sign-language/. [Accessed: 10 – Oct- 2022].

[5] P. Paudyal, A. Banerjee and S. K. S. Gupta, “SCEPTRE: a Pervasive, Non-Invasive, and ProgrammableGesture Recognition Technology,” Proceedings of the 21st International Conference on Intelligent User Interfaces, pp. 282-293, 2016.

[6] R. Y. Wang and J. Popovic, “Real-Time Hand-Tracking with a Color Glove,” ACM transactions on graphics (TOG), vol. 28, no. 3, 2009.

[7] R. Akmeliawati , M. P. L. Ooi and Y. C. Kuang, “Real-Time Malaysian Sign Language Translation using Colour Segmentation and Neural Network,” Instrumentation and Measurement Technology Conference Proceedings, 2007.

[8] F. S. Chen, C. M. Fu and C. L. Huang, "Hand gesture recognition using a real-time tracking method and hidden Markov models", Image and vision computing, vol. 21, pp. 745-758, 2003.

[9] M. A. Rahaman , M. Jasim, M. H. Ali and M. Hasanuzzaman, "Real Time Computer Vision-Based Bengali Sign Language Recognition," 17th International Conference on Computer and Information Technology (ICCIT), 2014.

[10] S. Padmavathi, M. S. Saipreethy and V. Valliammai, "Indian Sign Language Character Recognition using Neural Networks," IJCA Special Issue on Recent Trends in Pattern Recognition and Image Analysis RTPRIA, 2013.

[11] A. Chaudhary, J. L. Raheja and S. Raheja, "A Vision based Geometrical Method to find Fingers Positions in Real Time Hand Gesture Recognition," JSW, pp. 861-869, 2012.

[12] A. B. Jmaa, W. Mahdi, Y.B. Jemaa and A.B. Hamadou, "Hand Localization and Fingers Features Extraction: Application to Digit Recognition in Sign Language," International Conference on Intelligent Data Engineering and Automated Learning, pp. 151-159, 2009.

[13] G. Awad, J. Han and A. Sutherland, "A Unified System for Segmentation and Tracking of Face and Hands in Sign Language Recognition," 18th International Conference on Pattern Recognition, 2006.

[14] Z. H. Al-Tairi, R. W. Rahmat, M.I. Saripan and P.S. Sulaiman, "Skin Segmentation Using YUV and RGB Color Spaces," J Inf Process Syst, vol. 10, no. 2, pp. 283-299, June 2014.