International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

INDIAN SIGN LANGUAGE TRANSLATION FOR HARD-OF-HEARING AND HARD-OF-SPEAKING COMMUNITY.

Ashwith Poojary[1], Manasi Variar[2], Ritika Radhakrishnan[3]

Prof. Gayatri Hegde[4]

[1, 2, 3] UG Student, Dept. of Information Technology, Pillai College of Engineering, New Panvel, India.
[4] Faculty, Dept. of Information Technology, Pillai College of Engineering, New Panvel, India.

Abstract - *Sign language is an integral part of human communication as it has allowed people to communicate with the hard of speaking and hearing community and understand them better. However, not everyone is capable of using sign language which causes a barrier between. One finds it hard to communicate without an interpreter. With the help of deep learning and machine learning systems, we can eliminate said barriers. The purpose of our machine learning project is to create a web/phone camera based sign language recognition and translation system that would convert sign language gestures to text and vice versa in real time. It is possible to implement them via two ways : vision-based or glove-based systems. Capturing and translating the signs from the real life world will be the core objective of this project. Convolutional Neural Network (CNN) algorithm is used to implement our project. OpenCV video stream will be used to capture the real time gestures through the web camera or the phone camera. The preprocessed images are then fed to the Keras CNN model. We get the output in the form of text predicting the sign. Not only does each country have its own sign language but there are also many other regional sign languages too. Due to the Covid-19 pandemic, the alternative to normal communication is Video-calling, Facetime, etc. Hard-speaking and hearing people are not able to use such facilities effectively causing a hindrance in communication. Our paper aims to find a solution to such a problem and proposes a system for the translation of sign language using a webcam, mic, smart mobile phones, etc.*

Key Words: Hand movement, Gesture recognition, OpenCV, Sign language translation

1. INTRODUCTION

What is sign language? Sign language is a form of communication that makes use of a visual-manual medium to establish communication between two people without any sounds and spoken words. It is a visual medium of communication that is done with the help of hand signs, hand gestures, facial gestures expressions, and body language. It's the main form of communication for the aurally and speech challenged individuals i.e. Hard-of-Hearing/speaking community, There is no universal sign language, Sign languages are different around the world and even exhibit regional variants. 93% percent of the academic institutions here in India use Indian sign language [ISL]. Sign language has 3 major components that are distinguished on a visual basis : each finger-spelling, individual word-level sign vocabulary and non-manual(other) features. Here, the fingerspelling is used to spell the words by each individual letter. Lastly, the non-manual features such as mouth and body position, facial expressions and tongue, body langauge. In our project we focus on fingerspelling and try to perfect it. Examples of ISL are Twisting your mustache to convey the word 'man', and touching your nose to convey the word 'woman'. A sign language translator is a system that converts the given sign language symbols to text or voice in any native language. This type of sign language recognition and translator systems are called human-computer interaction systems. It is possible to implement them via two ways : vision-based or glove-based systems. For sign language translation systems we are using a vision-based system. The images are fed to the CNN model, a Deep Learning algorithm that can take in an input image. The output will be displayed as text for the hard hearing and aurally and speech-challenged person.

2. LITERATURE SURVEY

A. Sign Language to Text and Speech Translation in Real-Time Using Convolutional Neural Network:

Ankit Ojha et al. [1] presented a paper describing the use of various ML algorithms to translate sign language. In this paper, the OpenCV video stream is used to capture the gestures for American sign language through the camera. Hand gestures are taken for scanning, where preprocessing is done Keras. The model accumulates the recognized gesture to words. For detecting the gestures of the user, Convolutional Neural Network is being used. Kernel is used to scan all the pixels in the image. This project could have been built as a website. It has an accuracy of 95%.

B. Research of a Sign Language Translation System Based on Deep Learning:

The paper presented by Siming He [2] focuses on translating sign language into text using the newly developed R-CNN model. Faster R-Convolutional Neural Network model of Convolutional Neural Network is proposed, Faster R-CNN is a famous object detection architecture that uses convolution neural networks such as You Look Only Once and Single Shot Detector. Hand locating accuracy is greatly improved hence. The used data set is not diverse as all the sign language words are not included in the training-testing dataset. This paper has an accuracy of 99%.

C. Neural Sign Language Translation based on Human Keypoint Estimation.:

In this paper, authors Chang-Ko Jim and Sang-ki Ko [3] proposed a technique to detect sign language. We can utilize the human key points from hands, face, gestures and other body parts with the help of a neural network model for translating the videos of sign into natural language sentences. Normalization technique for which is used to preprocess the 2D coordinates of human key points is used in this project. The system is developed using OpenPose, which is an open source software. With the help of OpenPose, we can estimate and detect human keypoints in real time. The one disadvantage would be that technologies like AlphaPose, Mask R-CNN, etc are even more efficient as they perform better in terms of accuracy as compared to OpenPose. The translation model achieves an accuracy of 93.28% for 105 sentences.

D. An Integrated Two Way Indian Sign Language Translation System - A New Approach:

M.Suresh Anand et al. [4] conducted an analysis on the 2-way translation of the Indian Sign language into speech. In this paper, a webcam/ phone camera and microphone are used to implement the two way communication Indian Sign Language system. The system has two parts, translation of sign into a voice and reverse. The hand/fingers movements of an aurally and speech challenged person will be taken as continuous images from the input device like camera. The image taken from the camera will then be sent to various image preprocessing steps. The best frame amongst every taken input image will be taken into consideration. Feature Extraction which identities the proper meaning of the letters will take place. The recognized words are sent to the text from the voice converter to the receiver end.

Literature	Method	Accuracy
Ankit Ojha et al. 2020 [1]	CNN	95%
Siming He 2019 [2]	LSTM_fc 3D-CNN 3D ResNet-18+SVM_local 3D ResNet-18+SVM_fusion Our method_fusion	91.6% 91.5% 96.9% 98.3% 99.0%
Chang-Ko Jim and Sang-ki Ko. 2019 [3]	CNN, HMM	93.28%
M.Suresh Anand et al. 2013 [4]	Hidden Markov Model	-

Table -1: Summary of Related Work

3. PROPOSED WORK

Reading a person's sign gestures
1. Recognizing gestures
2. Preprocessing the gestures
3. Detecting the sign
4. Displaying the translated text of the sign

Recognizing a person's speech
1. Recognizing sentences
2. Filtering stemmed words
3. Text to sign translation

3.1 System Architecture

The system architecture is given in Figure 1. Each block is described in this Section.

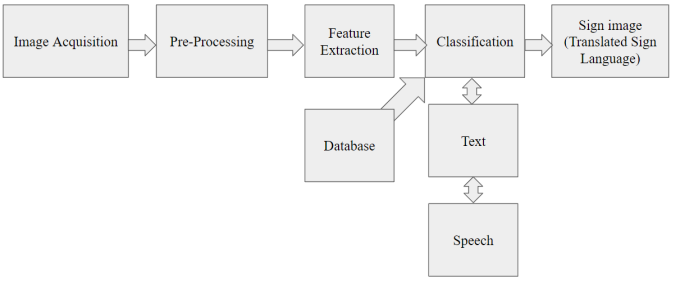

Fig. 1 Proposed system architecture

Sign to Text:

Image acquisition is the initial stage of our system. In this stage, we'll capture the images of the sign using the OpenCV video stream. The image will have an aspect ratio of 1:1. In Preprocessing, the image captured will be converted into a binary image. Other noises that could possibly contaminate the end result will also be removed in this stage. Feature Extraction is the stage where the image will then be converted into a grayscale image. Then in classification, the extracted feature is then compared to the images in the dataset, and based on its training level, it will provide the possible labels of the predicted signs. The label with the highest probability will be taken into consideration. Finally, the labels will be in a form of text, which will then be converted into speech using the pyttsx3 library. The end-user will receive the output in a form of text/speech.

Speech/Text to Sign:

The recorded audio from the user will be processed and all possible noise will be removed. The resulting output will then be converted into a text and will be stemmed accordingly. Based on the labels which the model received, it will check for all possible gesture translations and the gesture with the highest probability will be considered. This will generate a list of images of the gesture that will be displayed to the impaired user.

3.2 Requirement Analysis

The implementation detail is given in this section.

3.2.1 Software

The software components that will be used are;

Python 3.9.0, it is the most recent and updated version of the Python programming language, and it contains many new optimizations, libraries and features. Due to the amount of different types of libraries and modules available in Python, it is the go-to language for data science and machine learning.

OpenCV which is used in the development of computer vision applications is a cross-platform library which focuses on video capture,image processing, and analysis including features like face/object detection. It will take the data in real time as continuous images and process the same.

Jupyter Notebook, which is an open-source software that we will use to render all our code. Anaconda helps us to access jupyter notebook so that we can access the IDE and enter code.

Streamlit, a python based library used to design beautiful web apps which helps to integrate machine learning models into our UI. It provides a free hosting service to deploy our web app on the remote server.

3.2.1 Hardware

The paper makes use of web cameras for detecting the gestures of the user. The captured images were then sent for further processing. The output will be in the form of speech using speakers. Microphones are used to capture the user's speech for processing followed by translating it into sign images.

3.3 Dataset

Sign Language is a type of language that will use a set of movements and actions which when put together convey a meaning. These actions are predefined. Hard speaking and hard of hearing people usually use these languages to communicate with people. Not only the movement of hands but also the orientation of hands along with the combination of hand shapes are used to convey various signs. British Sign Language, Arabic Sign Language, Indian Sign Language, etc. are different types of sign languages. Indian Sign Language is the main sign language used in South Asian countries like Indian, Pakistan, Nepal. There are unique and distinguishing features of ISL that distinguish it from other types of Sign Languages. We created our own dataset using a python script that tracks and captures only the hand sign and processes each image. These images are then stored locally in a folder by the name of the sign.

4. RESULT

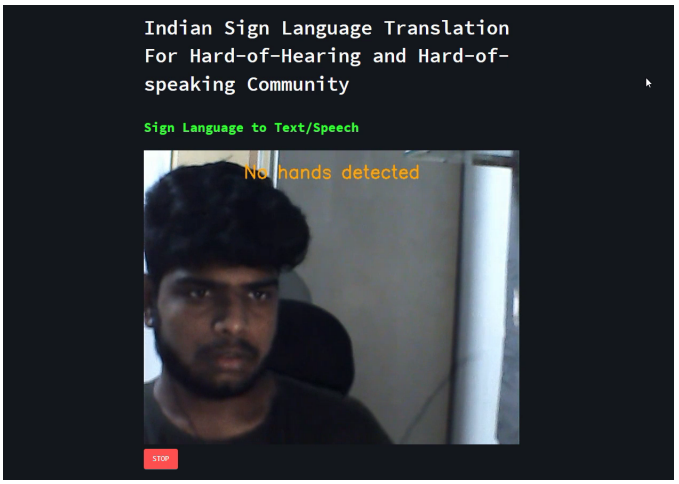

Fig. 2 Live video stream when the hands are not detected

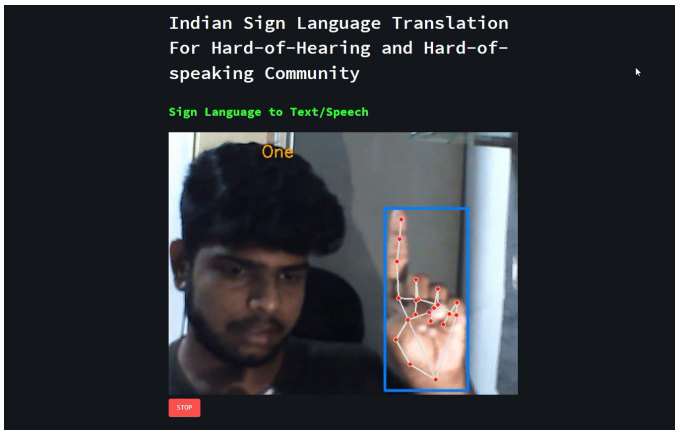

Fig. 3 Gesture Recognition for One

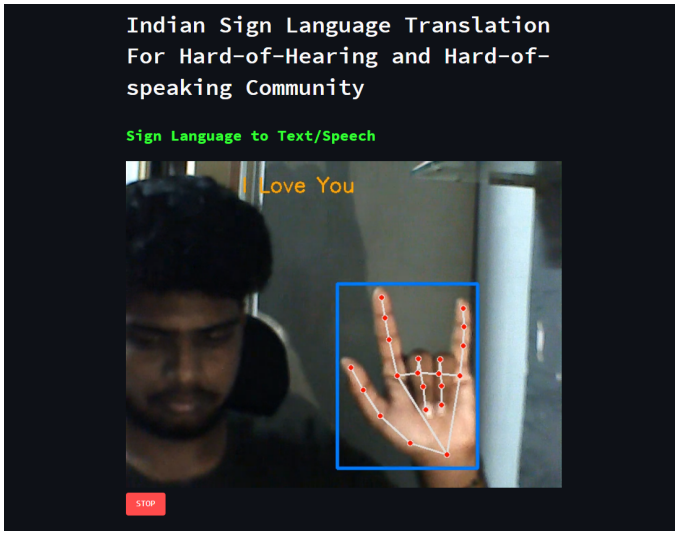

Fig. 4 Gesture recognition for I Love You

Fig. 5 Translating text into sign

Fig. 6 Translating speech to sign

5. CONCLUSIONS

Many technologies and ways have been emerging to help people who are hard at speaking and hearing, our project is one of them. When a proper sign recognition system is used then continuous communication can be done without any confusion. Our project as a whole is eventually just a demonstration of how Convolutional Neural Networks can be used to implement solutions to computer vision queries. Our paper helps in developing a detector and translator of the fingerspelling sign language The sign is taken in real-time via webcam and converted into text whereas text or speech can be taken via microphone and converted into a sign. A web app is created that integrates the machine learning model. The background needs to be dark and enough light is needed on the hands for the results to be as accurate as promised. Thus, the whole system comes together in a way that helps the hard speaking and hearing community for communication purposes.

ACKNOWLEDGEMENT

It is our privilege to express our sincerest regards to our Project Guide Prof. Gayatri Hegde for the valuable input, able guidance, encouragement, whole-hearted cooperation, and constructive criticism throughout the duration of this work. We deeply express our sincere thanks to our Head of the Department Dr. Satishkumar Verma and our Principal Dr. Sandeep M. Joshi for encouraging us and allowing us to present this work.

REFERENCES

[1] Ankit Ojha, Ayush Pandey, Shubham Maurya, "Sign Language to Text and Speech Translation in Real-Time Using Convolutional Neural Network", Volume-8 Issue-5, 2020.

[2] Siming He, "Research of a Sign Language Translation System Based on Deep Learning", Ridley College, Canada, Volume-1, 2019.

[3] Chang-Ko Jim, Sang-ki Ko, "Neural Sign Language Translation based on Human Keypoint Estimation", Volume-2, 2019.

[4] M.Suresh Anand, A.Kumaresan, Dr. N. Mohan Kumar, "An Integrated Two Way ISL (Indian Sign Language) Translation System - A New Approach", Volume-4, 2013.

[5] E. Abraham, A. Nayak, and A. Iqbal, "Real-Time Translation of Indian Sign Language using LSTM," Global Conference for Advancement in Technology (GCAT), BENGALURU, India, 2019.

[6] Bhargav Hegde, Dayananda P, Mahesh Hegde, Chetan C, "Deep Learning Technique for Detecting NSCLC", International Journal of Recent Technology and Engineering (IJRTE), Volume-8 Issue-3, September 2019.

[7] S. Shahriar et al., "Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning," TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018.

[8] Huang J, Zhou W, Zhang Q, et al. Video-based Sign Language Recognition without Temporal Segmentation[J], 2018.

[9] Umang Patel and Aarti G. Ambedkar, "Moment Based Sign Language Recognition for Indian Language", 2017, International Conference on Computing, Communication, Control and Automation (ICCUBEA).

[10] M.V, Beena. "Automatic Sign Language Finger Spelling Using Convolution Neural Network: Analysis.", 2017.

[11] M. Mahesh, A. Jayaprakash and M. Geetha, "Sign language translator for mobile platforms," 2017, International Conference on Advances in Computing, Communications, and Informatics (ICACCI), Udupi, 2017.

[12] Oyedotun, Oyebade K., and Adnan Khashman. "Deep learning in vision-based static hand gesture recognition."Neural Computing and Applications, 2017.

[13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[J].IEEE conference on computer vision and pattern recognition(CVPR) , 2016.

[14] L. Kau, W. Su, P. Yu, and S. Wei, "A real-time portable sign language translation system," 2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS), Fort Collins, CO, 2015.

[15] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553, 2015.

[16] Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. Sign Language Recognition Using Convolutional Neural Networks. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops, 2015.