JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ISL Translator for Caregivers

[1]Madhumita Menon, [2]Mayur Shinde, [3]Avinash Tripathy, [4]Gaurav Tawde

[1]Student, [2]Student, [3]Student,[4]Assistant Professor

[1-4]Electronics and Telecommunications Engineering,

[1-4]Vivekanand Education Society's Institute of Engineering, Mumbai, India

---

Abstract : Millions of people are deaf in India, let alone the entire world. The language that deaf people use to communicate with other people is quite uncommon and thus, the deaf community faces a lot of difficulty in communicating simple messages. The idea of educating the common masses about sign language has largely been futile and costly to begin with. Furthermore, it is quite a difficult task to enforce or create the awareness to learn sign language since the chances of a common person having an interaction with a deaf person is quite low. Hence, to solve this huge gap of communication that exists, we propose a sign language to text converter using LSTM based neural network. We focused largely on creating an ISL training database and attempting to teach certain words and phrases and recorded an accuracy of 94.22%.

IndexTerms – Computer Vision, Mediapipe Holistic, Tensorflow, Keras, Long Short-Term Memory, Indian Sign Language**.**

I. INTRODUCTION

The idea of being deaf or suffering from some sort of hearing disability has been a regularly ignored issue throughout the world. The WHO definition of "deafness" refers to the complete loss of hearing ability in one or two ears. The cases included in this category will be those having hearing loss more than 90 dB in better ear (profound impairment) or total loss of hearing in both the ears. The WHO definition of "hearing impairment" refers to both complete and partial loss of ability to hear. Going by this definition, over 5% of the entire global population or roughly 350 million people suffer from some sort of hearing disability.

According to WHO statistics, every 4th person in the world would suffer from hearing loss by 2050.[7] People who have hearing loss use sign language to communicate with other people. In India alone, the WHO estimates that there are 63 million people who suffer from Significant Auditory Impairment. This places the disability at a staggering 6.3% of the Indian population. It is estimated that by 2050, every 1 in 4 children in India will suffer from some form of hearing disability. To summarise, Hearing Disability is a bigger problem than Vision Impairment.

To solve this problem, we have Sign Language which is creating a combination of signs using your hands to indicate either words or numbers and sometimes even phrases. Every country has its own version of sign language. For example, the USA has the American Sign Language or the ASL while India has the Indian Sign Language or ISL. Learning basic letters and numbers is quite easy and can be learnt within a few hours. But the problem arises in trying to gain mastery over Sign Language. To master any language, one needs to have proper knowledge of its grammar as well as the syntax and that is where the main problem arises. To learn Sign Language, you need to invest time that can vary from 6 months to even 2 years depending on what level of mastery one wishes to acquire and seeing the application of sign language to be so limited and rare, it doesn't seem to be quite beneficial for people to invest their time in. To add on to this, the learning curve of sign languages differ from country to country and since every country has its own version of it, if one is supposed to travel to a different country, their efforts of learning sign language are futile since they would have to start from scratch again.

The Indian Government did try to address the problem by introducing the The Persons with Disabilities (Equal Opportunities, Protection of Rights and Full Participation) Act 1995. The Act provides for both the preventive and promotional aspects of rehabilitation such as education, employment, vocational training, reservation, research and manpower development, creation of a barrier- free environment, unemployment allowances and special insurance schemes for disabled employees, establishment of homes for persons with severe disability, and so on. The lack of funds over the years have plagued the efforts that were so greatly cheered on by the Deaf Community of India and as we speak today, the developments have come to a standstill just due to the lack of infusion of funds.

Seeing so many problems arise when going by the traditional method, we have attempted to solve this problem using the technology available to us. Our solution revolves around a concept called the Long Short-Term Memory, RNN and Machine Learning. Our prototype has the ability to scan the sign shown by a person and convert it into text or speech within a few seconds. Our prototype can train itself over time through the data that goes through it every time it converts sign language to text. The

current accuracy of our model in correctly converting sign language to text is 94.2 % and this figure will only increase as we feed it more and more data.

Our proposed solution primarily aims at providing a cost-efficient method to caretakers of elderly people. It has been observed that finding skilled individuals for taking care of the elderly while possessing the knowledge of sign language is a time intensive job and implementation of our system has the potential to save both time and money. Apart from caretakers, our system eliminates the need to hire a professional translator for people with hearing disability as can be seen as a very common practice for the deaf community when travelling someplace far.
Existing solutions provide relatively low accuracy or have long processing time.

II. RELATED WORKS

There has been considerable work in the field of Sign Language recognition with novel approaches towards gesture recognition. Different methods such as the use of a 2D camera or Microsoft Kinect Sensor have been employed earlier. A study of many different existing systems has been done to design a system that is efficient and robust than the rest.

Muthu Mariappan H and Dr Gomathi V came up with a model where a camera unit captures gestures of people who suffer from hearing and speech impairments. The raw videos can be taken in any sort of background and fed as input to the system. The image frames are resized as a precautionary measure to maintain an equality among all the videos. OpenCV which is short for Open-Source Library for Computer Vision is used for feature extraction and video classification. A data sample of 80 words and 50 sentences were recorded from 10 volunteers or 800 words and 500 sentences data set was fed to the system for data training and the testing returned an accuracy of 75% for gesture recognition.[1]

Kartik Shenoy, Tejas Dastane, Varun Rao, Devendra Vyavaharkar have proposed a model wherein it will recognise hand poses and gestures from the Indian Sign Language using Grid-based features. It has the ability to identify 33 hand poses and some gestures from ISL. The model captures hand poses using an Android smartphone camera and transmits the image of that pose to a remote server for processing. The image is subjected to a grid-based image feature extraction technique which represents the hand's pose in the feature vector. The hand poses are classified using k-nearest neighbours' algorithm while gesture recognition is done using Hidden Markov Model Chains. The model has a pre-processing phase which includes face removal, stabilisation and skin colour segmentation to remove background details and reduce noise. For the recognition of hand poses, features are fed into a classifier. Recognised hand pose is sent back to the android device. A dataset of close to 25,000 images helps this model in achieving an accuracy of 99.7% for static hand poses and 97.23% for gesture recognition.[2]

Tülay KarayÕlan and Özkan KÕlÕç proposed a system which converts sign language to text by an automated sign language recognition system based on Machine Learning. The proposed system follows a 3-step process to convert sign language. It starts with image processing. The image can either be in the local system or the system can take input from a webcam camera. After processing the input image, the classifier classifies the image according to the class it belongs to. The system has 2 classifiers: one uses raw image features and the other one uses histogram features. These classifiers use Backpropagation Algorithm which is a commonly used Artificial Neural Network learning technique used to classify images. First classifier which is called the Raw Features Classifier uses 3072 features while the second classifier which is called Histogram Features Classifier uses 512 features. Marcel Static Hand Posture Database was used to train the system and the system gave an accuracy of 70% for the Raw Feature Classifier and 85% accuracy rate for the Histogram Features Classifier.[3]

Dan Guo, Wengang Zhou, Houqiang Li, Meng Wang proposed a model wherein they use Hierarchical Long Short-Term Memory for Sign Language Translation. The unique part about this model is using the Hierarchical LSTM to encode the visual semantics of the sign. A 3D CNN model is used to extract visual features as compared to the 2D CNN in the earlier models. 3D CNN helps in better understanding of spatio-temporal context and also helps in avoiding dependency of the long sequence transmission in the LSTM learning. The signs are easily recognised by discriminative gestures by utilising an online adaptive key clip mining method by optimising residual square sum of previous successive 3D CNN features and current feature, and capture their linear correlation. The main motivation behind doing all this is to avoid training the model with less-important clips that may degrade the performance as well as accuracy instead of improving it.[4]

Anshul Mittal, Pradeep Kumar, Partha Pratim Roy, Raman Balasubramanian and Bidyut B. Chaudhuri proposed a model for tracking and recognising continuous ISL sentences. A modified LSTM classifier is used for the recognition of continuous signed sentences using sign sub-units. The Convolutional Neural Network (CNN) is used to extract the spatial features from the signed sequences which are modelled by modified LSTM for recognition. A dataset of close to 3150 sign words and 157 sign sentences was used to train the model and an accuracy of 89.5% was recorded on the sign words while the recognition of sign sentences gave an accuracy of 72.3%. [5]

III. PROPOSED METHODOLOGY

This section will explain our LSTM based neural network architecture that we have used for our sign language translator.

3.1 Data Collection

Most of the words in sign language are dynamic, that is they are represented by an action or a series of actions. Hence, we made videos of 19 words in Indian Sign Language for our data set. The videos are of waist-length performing a gesture of the Indian Sign Language word with their arms bent at the elbow approximately at chest level. To make the model predict accurate results irrespective of camera quality, lighting and other external factors, we decided to take help from friends and family by

having each of them make videos for those same 19 words. This way we had different types of videos for each word. For example, one of the videos was recorded in dim lighting, some of the participants were in an outdoor setting, there were participants from different ages, genders, etc.
Additionally, to increase the sample size of the dataset, a large number of videos were also captured using the webcam.

Table 1: Words

Allergy	Hospital
Ambulance	Hungry
Appointment	Hurt
Blood	Medicine
Call	Sick
Calm	Thermometer
Doctor	Toilet
Feel	Water
Food	Where
Help	

3.2 Pre-processing the video dataset

Each video is converted to .mp4 extension and the frame rate is set at 25 fps. Some videos were kept at a lower quality so that users with a relatively weaker hardware support may also avail this webapp.

3.3 Feature extraction using Mediapipe holistic

This solution is available for working with both a live video stream from a webcam and photo and video files. In the case of Python, MediaPipe is available as a Python module package. Mediapipe Holistic, determines key landmarks and stores them in a numpy array. So, at the end of the feature extraction process, we have a .npy file for each frame and a set of .npy files for each video.

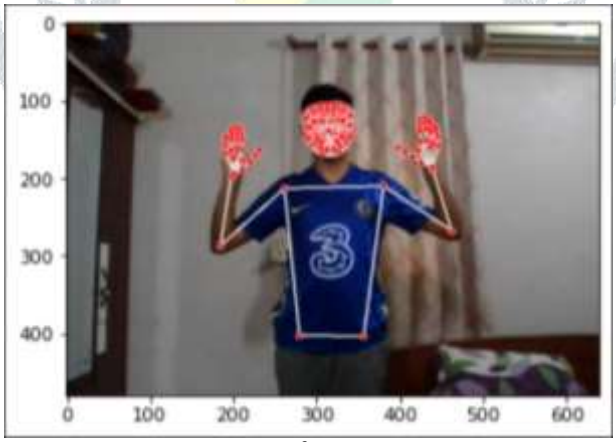

Fig 1: feature extraction

3.4 Long Short-Term Memory (LSTM) Network

LSTM is an improvement from the traditional RNN. RNN is basically efficient only while dealing with short-term dependencies. It does not have memory, meaning it does not remember what happened before this. This issue was resolved by slightly tweaking it, which gave birth to: Long Short-Term Network's make small modifications to the information by additions and multiplications. Therefore, it can selectively retain or discard information.

3.5 Web App

A user interface is very essential so that everybody is able to access the required resources. Our Web App will access the front camera, web camera or a stand-alone camera connected to a computational device for capturing a video feed input. It will simultaneously display the actions being signed by the user. In this process, the time lag is next to nothing and hence translation occurs in real time.

3.6 Algorithm

i. The participant signs into a camera that captures a video feed subsequently collecting the frames.
ii. The key points or landmarks are then collected and sent to the model.
iii. The model then classifies the frame and accordingly the word is predicted and displayed on the screen.

IV. RESULTS

This user-friendly web application is a conversational tool that translates ISL for challenged people. The product specifically targets medical professionals who take care of speech impaired patients and have very limited knowledge of Indian Sign Language. Therefore, the number of words is kept at a minimum and is highly focused for this very purpose. This means that all 19 words and other words that will be added in the future will cater to the medical needs of the speech impaired. The limited database of words in turn helps with increasing the accuracy of the application and reduces the computational time. Thus, the final product will help patients convey their feelings accurately to their caretakers.

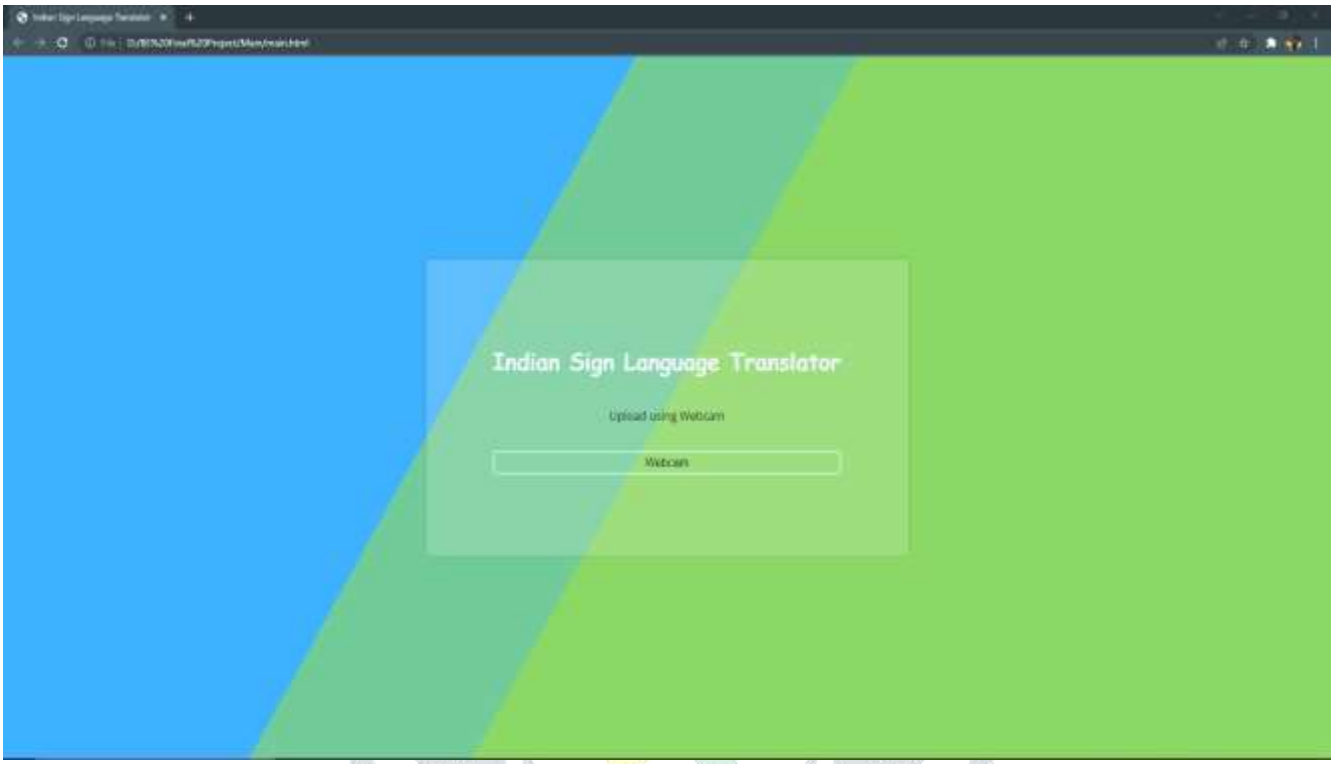

Fig 2: Webapp

Fig 3**:** Model successfully recognizing "Allergy"

Fig 4**:** Model successfully recognizing "Ambulance"

Fig 5**:** Model successfully recognizing

"Appointment"

IV. FUTURE SCOPE

The accuracy of the model can be further improved by increasing the dataset. The application can be extended to a sentence level. Work is underway for text to sign translation and this could be integrated into the same web application. Finally, the number of words and videos can be increased and this could be achieved by putting up a feature for the existing users to contribute towards our dataset.

V. CONCLUSION

The final product successfully translates signs to text on a word basis. The aim of this project, that is to help families who cannot afford caretakers with the knowledge of Indian Sign Language can possibly be achieved. Moreover, the resources for Indian Sign Language are relatively scarce, hence we could collect our own database for this project. The model works well and gives an accuracy of 94.22%.

VI. ACKNOWLEDGEMENTS

We would like to extend our sincere gratitude to our project guide Prof. Gaurav Tawade for guiding us through with completing this project successfully.

REFERENCES

[1] M. Mariappan, H; Gomathi, V (2019). [IEEE 2019 International Conference on Computational Intelligence in Data Science (ICCIDS) - Chennai, India (2019.2.21-2019.2.23)] 2019 International Conference on Computational Intelligence in Data Science (ICCIDS) - Real-Time Recognition of Indian Sign Language. , (), 1–6. doi:10.1109/ICCIDS.2019.8862125

[2] K. Shenoy, T. Dastane, V. Rao and D. Vyavaharkar, "Real-time Indian Sign Language (ISL) Recognition," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2018, pp. 1-9, doi: 10.1109/ICCCNT.2018.8493808.

[3] Tülay KarayÕlan and Özkan KÖlÕç, "Sign Language Recognition", IEEE, 2017

[4] D. Guo, W. Zhou, H. Li, and M. Wang, "Hierarchical LSTM for Sign Language Translation", AAAI, vol. 32, no. 1, Apr. 2018.

[5] A. Mittal, P. Kumar, P. P. Roy, R. Balasubramanian and B. B. Chaudhuri, "A Modified LSTM Model for Continuous Sign Language Recognition Using Leap Motion," in IEEE Sensors Journal, vol. 19, no. 16, pp. 7056-7063, 15 Aug.15, 2019, doi: 10.1109/JSEN.2019.2909837.

[6] Varshney S. Deafness in India. Indian J Otol 2016;22:73-6

[7] https://www.who.int/en/news-room/fact-sheets/detail/deafness-and-hearing-loss

[8] https://github.com/google/mediapipe

[9]https://indiansignlanguage.org/?__cf_chl_rt_tk=r3irIrEYzYfnBXBzgc1hsyTvnmDaCPiJHOT8XZSSjuU-1646455360-0-gaNycGzNCOU

[10] https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/