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

REAL TIME SIGN LANGUAGE RECOGNITION

[1]Asmita Das, [2]Pradeep Shastry K S, [3]B Vedanth, [4]Gaurav R, [5]Niroshan P, [6]Dr. M.S Nidhya

[1,2,3,4,5]Student, [6]Associate Professor
[1,2,23,4,5,6]School of CS and IT
Jain (Deemed-to-be-University), Bengaluru, Karnataka

Abstract : **This research describes an innovative technique for communicating with people who have speaking and hearing impediments. It discusses a new approach for sign detection and converting signs to text. The developed system can extract indications from video sequences with a minimally congested and dynamic background. It recognizes static and dynamic gestures and extracts the appropriate feature vector for each. Action Recognition using Python is used to classify them. The findings of the experiments show that signs can be segmented satisfactorily against a variety of backdrops, and that gesture and speech recognition can be recognized with reasonable accuracy.**

Keywords: sign-language, sign-language detection, ASL, LSTM, MediaPipe.

I. INTRODUCTION

It's tough to speak with people who have a hearing impairment. Deaf and mute persons communicate via hand gesture sign language, that makes it difficult for non-deaf and mute folks to know their language. As a result, technologies that recognize numerous indicators and communicate the knowledge to normal people are required [4].
Sign language is a visual language. It mainly consists of 3 major components [1]:

A. *Fingerspelling*: Words are spelled out letter by letter, and use hand gestures to indicate the meaning of words at the word level. This is accomplished using the static Image Dataset [1].
B. *World-level sign vocabulary*: Video categorization recognizes the whole words or letters. (Dynamic Input / Video Classification) [1].
C. *Non-manual features*: Facial expressions, tongue, mouth, body positions.

The development of a real-time sign language translator is a significant step forward in improving communication between the deaf population and the general public. The creation and execution of an American Sign Language (ASL) fingerspelling translator are presented here. The deaf community benefits greatly from the use of American Sign Language (ASL). However, there are not many speakers, limiting the number of persons with whom they are able to interact comfortably. When an emergency comes, textual communication is inconvenient, impersonal, and even impractical. We describe an ASL recognition method to help overcome this barrier and enable dynamic communication. (Shah, December 2018

II. LITERATURE REVIEW

A. *MotionSavvy UNI*: Price: $198 USD (₹14,854 INR) .
A two-way communication device that allows the deaf and the hearing to converse with one another. UNI is a smartphone app that watches your hands in real time and converts your gestures into spoken English [12].

B. *Voicer:* It is a programme that helps deaf people communicate more easily on a daily basis. This application's main role is to detect and interpret American Sign Language (ASL). Electromyography (EMG) sensor technology has advanced considerably in recent years. EMG sensors can accurately detect movements of the arms and even the fingertips. This opens up a new avenue for deaf people to solve their communication issues. The development of EMG sensing technologies also influenced the design of Voicer.

III. PROPOSED METHODOLOGY

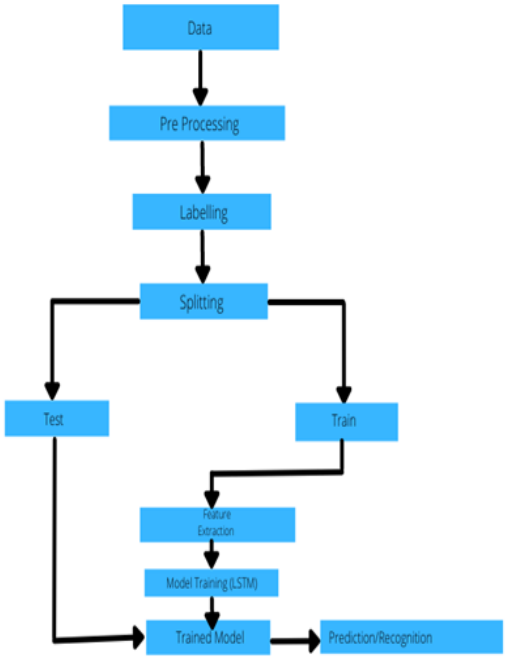

fig. 1 representation of proposed methodology as a flowchart.

The dataset is first collected and fed into the system for pre-processing.

Then, we are processing the dataset to make it suitable for our model. The key points are extracted and stored as NumPy arrays. In the collected dataset, the color format of images does not match the color format expected by the MediaPipe framework. It is therefore changed. After that, the extracted NumPy arrays are concatenated with their respective labels (/words).

The dataset is then split into the training set and the testing set. We've used 30% for testing and 70% for training purpose [9]. The training dataset is passed through feature extraction to only extract the required data. The processed training dataset is then trained using sequential LSTM layers along with Dense layers. The weights and biases are stored [9].The test dataset is passed to the trained model to check whether the model is working or not. Finally, the trained and stored model is loaded to make real time predictions by capturing video via webcam.

Some of the algorithms used by us are as follows:

A. MediaPipe Holistic: Separate models for stance, face, and hand components are integrated into the MediaPipe Holistic pipeline, each of which is optimised for its own domain. The input to one component, however, is not well-suited for the others due to their varied specialties. For instance, the pose estimation model uses a smaller, fixed

resolution video frame (256x256) as input. The image quality would be too low for proper articulation if the hand and facial sections of that image were cropped to pass to their respective models. As a result, we created MediaPipe Holistic as a multi-stage pipeline that processes each zone with the optimal image resolution for that location.

B. *Dense Model:* The dense layer is a deep-connected neural network layer, meaning that each neuron in the dense layer receives input from all neurons in the previous layer. In the model, we can see that the dense layers are the most used layers. The dense layer performs matrix-vector multiplication in the background.

C. *LSTM:* Long Short-Term Memory Network is an advanced RNN that is a sequential network that enables information persistence [2]. You can address the vanishing gradient problem that RNNs face. Recurrent neural networks, also known as RNNs, are used for persistent storage [2]. Suppose you remember the previous scene when you watch a video, or you know what happened in the previous chapter when you read a book. RNNs work as well, remembering previous information and using it to process the current input. The downside of RNNs is that they can't remember long-term dependencies because of the disappearance of the gradient. LSTM is explicitly designed to avoid long-term dependency issues [2]

IV. DATA FLOW DIAGRAM

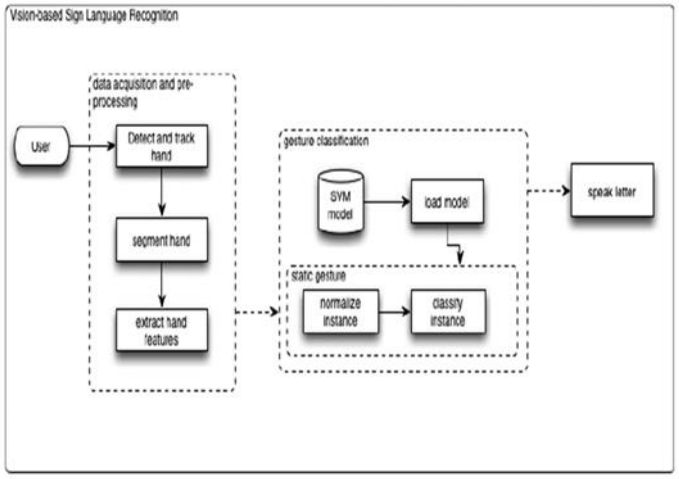

fig. 2 basic vision-based sign-language recognition dfd.

In the first steps, the running model is waiting for an action or a posture.
Once the pose is read by the webcam, a video sequence of images is transferred to the MediaPipe model in real-time. After that, the MediaPipe model identifies and collects data from the live stream, tracking the poser's hand and extracting its dimensional properties as well as facial features.
The extracted data from the previous model is provided in real-time to the trained model to assess and categories the instance if it has a valid pose according to the training. Otherwise, the instance is normalized.
As seen in Fig 3.3, the categorized instance is subsequently returned to the poser in running text format.

V.OUTPUT SCREENSHOTS

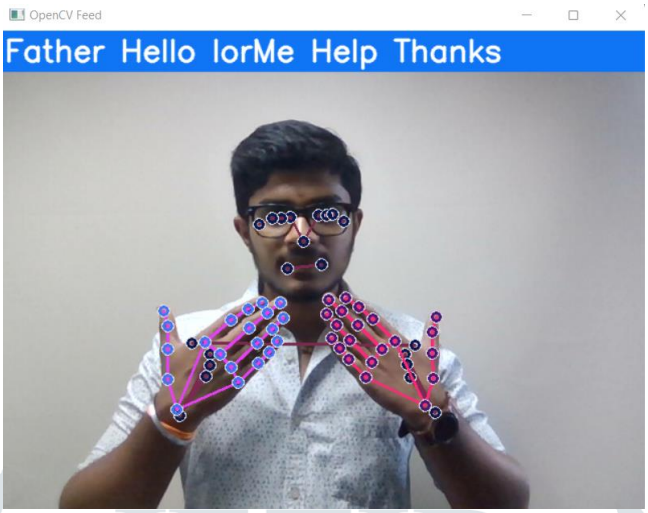

fig. 3.1 output screenshot of the model recognizing the word ‘thanks’.

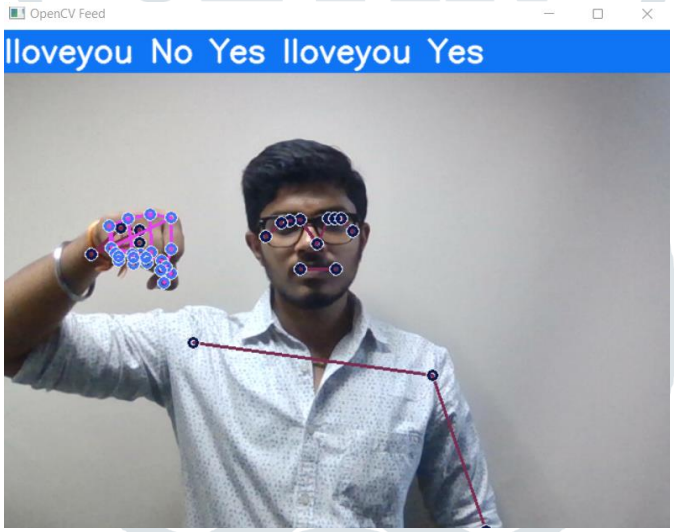

fig. 3.2 output screenshot of the model recognizing the word ‘yes’.

fig. 3.3 output screenshot of the model recognizing the word ‘mother’.

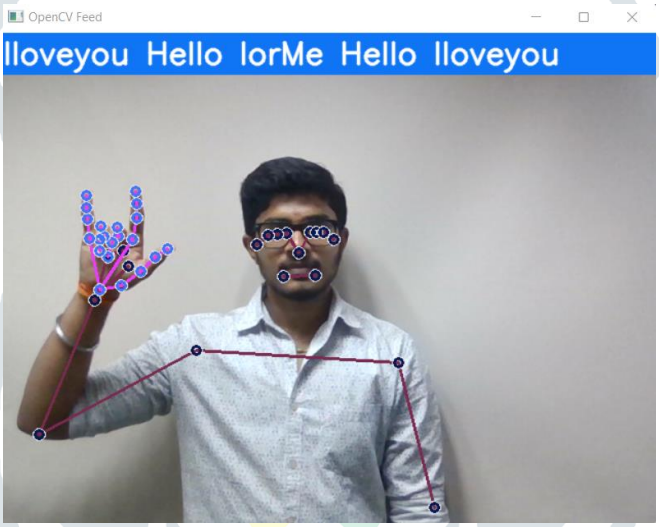

fig. 3.4 output screenshot of the model recognizing the word ‘iloveyou’.

VI.CONCLUSION

This research paper primarily focuses on solving two-way communication problem for the deaf and mute community. Most of the social problems for the community stems from the lack of effective communication. This presents a need for a medium of communication, which can translate the sign language used into text which can be understood by people who do not know sign-language. Through our paper we have established a medium that does the same.

There are areas in which our paper can be further improved. Our paper currently can only recognize poses done by one person in the frame. There are several sign languages being used around the world. Our model currently only recognizes the most widely used sign language, which is ASL (American Sign Language). Currently, our paper can only identify a limited set of words. We plan on extending our library as we go forward.

REFERENCES

[1] Anirudh Tunga, S. V. (2021). Pose-based Sign Language Recognition using GCN and BERT. Purdue University, Department of Computer Science. IEEE Xplore.

[2] Dongxu Li, C. R. (21 Jan 2020). Word-level Deep Sign Language Recognition from Video. The Australian National University. arXiv:1910.11006v2 [cs.CV].

[3] Ming Jin Cheok, Z. O. (Received: 23 June 2016 / Accepted: 31 July 2017). A review of hand gesture and sign language recognition techniques. © Springer-Verlag GmbH Germany 2017.

[4] Shah, J. (December 2018). A DEEP-LEARNING ARCHITECTURE FOR SIGN LANGUAGE RECOGNITION. THE UNIVERSITY OF TEXAS AT ARLINGTON, Department of Computer Science.

[5] Agarap, F. A. (2017). An Architecture Combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for Image Classification. ArXiv.

[6] Herath, H.C.M. & W.A.L.V. Kumari, & Senevirathne, W.A.P.B & Dissanayake, Maheshi. (2013). "IMAGE BASED SIGN LANGUAGE RECOGNITIONSYSTEM FOR SINHALA SIGN LANGUAGE".

[7] M. Geetha and U. C. Manjusha, (2012) "A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation(IJCSE), vol. 4, no. 3, pp. 406-415. 2012.

[8] Pigou, L., Dieleman, S., Kindermans, PJ., Schrauwen, B. (2015). Sign Language Recognition Using Convolutional Neural Networks. In: Agapito, L., Bronstein, M., Rother, C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science(), vol 8925. Springer, Cham. https://doi.org/10.1007/978-3-319-16178-5_40

[9] Kexin Zhang, (2016) Research School of Computer Science, Australian National University "LSTM: An Image Classification Model Based on Fashion-MNIST Dataset" U6342657@anu.edu.au

[10] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. (2017) "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields."In CVPR, 2017. 5

[11] X. Chu, W. Ouyang, H. Li, and X. Wang. (2016) "Structured feature learning for pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,"pages 4715–4723, 2016. 6

[12] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. (2015) "Long-term recurrent convolutional networks for visual recognition and description."pages 2625–2634, 2015.

[13] V. Athitsos, C. Neidle, S. Sclaroff, J. Nash, A. Stefan, Q. Yuan, and A. Thangali. (2008)"The American sign language lexicon video dataset". IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pages 18.

[14] P. C. Badhe and V. Kulkarni. (2015) "Indian sign language translator using gesture recognition algorithm." In 2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS), pages 195–200.

[15] Muskan Dhiman (2017) National Institute of Technology, Hamirpur (H.P.) "*SIGN LANGUAGE RECOGNITION"* Summer Research Fellowship Programme of India's Science Academies 2017.