Indian Sign Language Communication

Dhiraj Jadhav[1], Balaji Padamwar[2], Shriram Pareek[3],

Kishan Partani[4],Md Mubasheeruddin Siddiqui[5]

[1]*Professor, Computer Engineering, Vishwakarma Institute of Technology, Pune.* dhiraj.jadhav@vit.edu

[2]**Students,** *Computer Engineering, Vishwakarma Institute of Technology, Pune.*
balaji.padmawar18@vit.edu, shriram.pareek18@vit.edu

kishan.partani18@vit.edu, mubasheeruddin.siddiqui18@vit.edu

Abstract: Sign languages are the primary source of communication among dumb as well as deaf people and no proper structure is present around the same, also increasing technology has not contributed as per its potential across this domain, Lot of research is done in this area focusing on American Sign Language(ASL) and hence adequate data is available for the same but when it comes to Indian Sign Language(ISL) it's a bit lagging behind. There are tools available for ASL but no tool is present for Indian Sign Language Community, hence here we implement a word level ISL tool which uses pose and gestures for recognition and also a dictionary for easy finding of actions of unknown words.

Keywords - Indian Sign Language (ISL), Action Recognition, Pose Estimation, Neural Networks, Human Intervention.

I. Introduction

Deaf and dumb people are a major part of our society comprising of around 5% of population in the world [12]. The reports (sept. 2018) by World Health Organisation (WHO) says that nearly 63 million people in India suffer from either partial or complete deafness [8]. Out of this population 50 lakh are children.

There are nearly 300 sign languages being used in the world. Sign language represents English alphabets using finger spelling [4]. According to the report, more than 70 million deaf people around the world use sign language for communication. In the Indian subcontinent, ISL is the predominant language. As of 2021, ISL is the most used sign language in the world. ISL follows a 2 handed style to denote a gesture.

Irrespective of being so widely used language, ISL unlike other popular languages has no proper resources available currently [4] and there have been appreciable efforts to change the same and recently ISL dictionary has been launched which is the great start and contribution towards the community [9]. And our project aims to contribute in the same direction. Some of the words and the respective gestures are shown in fig 1.

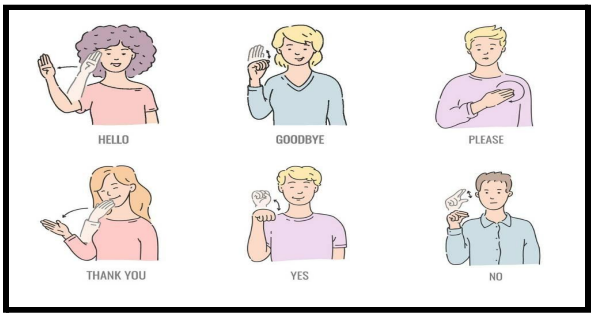

Fig.1 Sample Sign Language [10]

The motive behind the project is identifying word level signs in ISL from the corresponding gestures. An AI based model is developed that uses gestures and through pose estimation it

extracts coordinates across various parts of the body, normalizes them and then are passed across a Long Short Term Memory (LSTM) based model.

II. Literature review

Li et al. described the use of a large scale Word-Level American Sign Language (WL-ASL) dataset which covers a broad range of words for evaluation of various deep learning models. Various baseline architectures such as 2D convolutional Recurrent Neural Network (RNN), 3D Convolutional Neural Network (CNN) [1,6], Pose RNN, Time domain Graph Convolutional Network (TGCN) are implemented and compared. They realized that sign language requires a specific domain knowledge such as everyday terms based on agriculture, technology, etc. which makes labeling large amounts of samples per class unaffordable. They aim to implement advanced few-shot learning for facilitating sentence-level as well as story level machine sign translations [1]. Working of those models is as shown in fig 2 and fig 3.

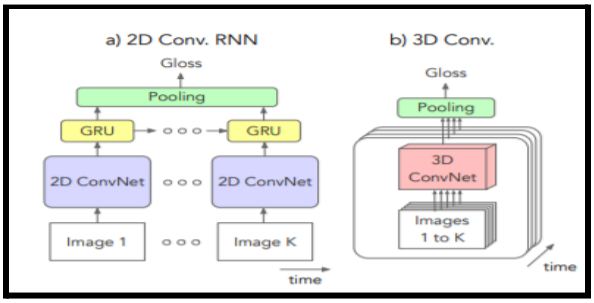

Fig.2 Baseline Architecture I [1]

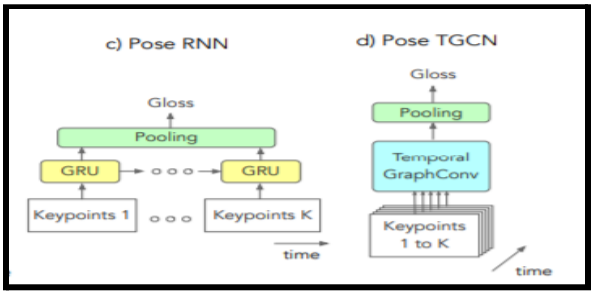

Fig.3 Baseline Architecture II [1]

Khan et al. had used various methods such as Neural Networks [1], Hidden Markov Model (HMM), fuzzy c-means clustering can be used for the gesture recognition [2]. The process of gesture recognition is divided into 3 parts i.e. extraction method, feature estimation, and classification [2]. HMM tools are perfect for recognizing dynamic gestures [2] but are more computational or consume more resources. Recognition process is affected by the proper selection of feature parameters and suitable classification approach. Different approaches may lead to different accuracies in recognition. Also the lighting conditions affect the performance of the model [2].

Tripathi et al. focused on varied features such as speed of gestures, shape of the hands while acting or forming gestures etc [3]. Gesture recognition system is wholly dependent on key frame extraction for the further processing. Even loss of one important frame can cause change in the gesture recognition. Proper extraction can help get the needed number of frames for the respective gesture recognition. Considering the distance based classifiers, the performance of euclidean distance and correlation is better than the others like city-block distance, chessboard distance, etc. in terms of recognition rate[3].

Shirbhate et al. presented an automatic sign language recognition system in real time using different Machine Learning (ML) tools such as Hand Segmentation, Random Forest, Support Vector Machines (SVM). Here learning and classification is carried out after extracting relevant features from skin segmented images. But the system is designed to be used only for static ISL numeral signs. Also they feel that the ISL lags behind its counterpart ASL due to lack of standard datasets. ISL uses both hands for gestures unlike ASL which results in obstruction of features. Also some characters share the same

alphabet (such as W and 3, they have the same sign) and resolving such characters is context dependent [4].

Number of techniques are available for gesture recognition even from using hardware such as gloves to softwares such as openCV. Furthermore techniques consist of a color based approach that provides a contactless communication between humans and computers. However this involves various challenges including lighting variation, background noise, complex background and a lot more. This method can be further extended to skin color approach or using gloves of a specific color [5]. Another technique that comes up in the discussion is motion based recognition that tracks down the motion of the body at the time of gestures and based on the moment it classifies or predicts the correct word regarding it. Next on the list is depth based approach, that uses different types of cameras providing a 3D geometric information about the object. 3D model based recognition [1, 6] and Deep learning based recognition also made it to the list that make use of neural networks for gesture recognition.

Hakim et al. solved the problem of real time gesture recognition by making the varied combinations of RGB colors and depth modalities that acts as an input to the deep learning model [1, 6]. The model could extract the coordinates or the positional features of sequence of gestures, particularly for dynamic gesture recognition. They aimed to use transfer learning techniques for future changes by training the model on large datasets such as the Sports1M dataset. [6].

Sinha et al. tracked down the motion of the gestures rather than just focusing on the frame of action [7]. MediaPipe was used to get the position or the coordinates of the gestures as shown in fig 4. The MediaPipe module collects data from the camera, processes it and generates key points. Also the model for gesture recogniser comprised 2 neural networks one for the static gestures and another for the dynamic gestures [7]. Optimization is done to segregate the frames between static and dynamic gestures. This helped it to achieve high detection accuracy for multiple users and in different lighting conditions.

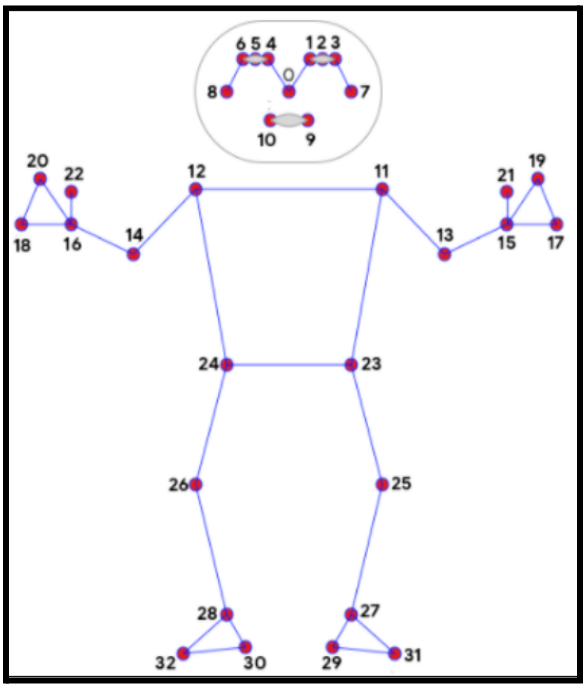

Fig.4 MediaPipe - Pose Landmark Model [11]

Based on the related work mentioned above and the conclusions gathered from it, gesture recognition by motion detection using MediaPipe has an edge over other techniques. Neural Networks have played a vital role in providing a greater accuracy but on the other hand Neural Networks on a single frame at a time may not be that beneficial as a gesture may consist of both static or the dynamic action/word. Hence for gesture recognition all those key frames need to be considered. Hence we have used LSTM and MediaPipe for the same and have discussed in the further section about their usage in detail.

III. Methodology

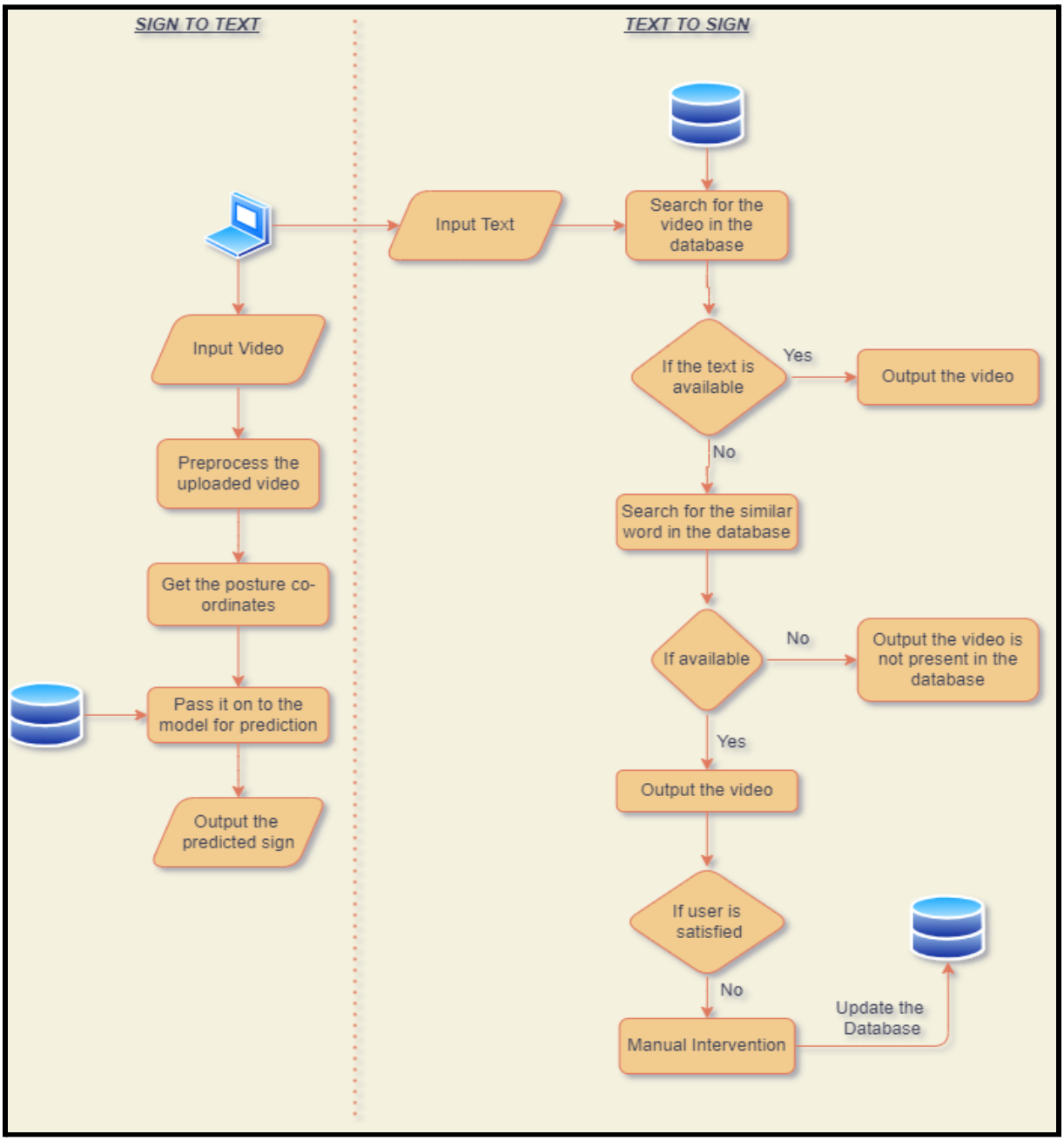

Fig.5 System Architecture Diagram

The System Architecture diagram in fig 5 depicts the flow of our project. It states how the 2 ends i.e. Text to Gesture and Gesture to Text exactly works. And the algorithm for the same is stated below.

1. Gesture to Text

Algorithm for Gesture to text comprises of following steps :

1. Users can upload a video or can act in front of a webcam and then frames would be collected from it.
2. Then the necessary points or positions of the user are gathered and then these are passed on to to the model.
3. Based on the model the video or the frames are classified and hence a word is predicted.

1. Text to Gesture

Text to gesture algorithm comprises of steps shown below :

1. Users can enter a text as an input to fetch a particular gesture.
2. The word then is preprocessed and then searched in the database and if the word is present then the specific video is returned
3. Else a similarity of words is checked then the video with max similarity is fetched and if the score crosses a certain threshold then that video is returned.

IV. RESULTS AND DISCUSSIONS

This is a great conversation tool based on ISL. The final product is a web application that successfully recognises the gesture and predicts the corresponding text for the same. So that a sign or a gesture by a challenged person (dumb person) can be understood by the normal people. This is done with great accuracy and is also user friendly that can be easily handled by anyone. The example of the same i.e. Gesture to Text depicting the action of 'Technology' is shown in fig 6.

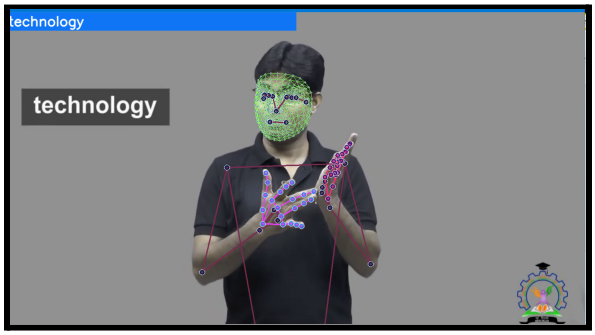

Fig.6 An example of Gesture to Text

On the other hand it also suggests an action for the given word. This will help normal people to reply back to the ones who won't have the ability to listen (deaf). As this end also takes care of the relative words hence using similarity calculation this makes it easy for the user to get the desired action even if the actual word is not present in the database. On the top of that, users can also contribute to increasing the dataset as they require by raising a concern if the required word is not present in the dataset. Hence the users could take care of their own dictionary of actions that they use in their day to day life. Example of the Text to Gesture comprising the word 'Canada' and it's relevant action is shown in fig 7.

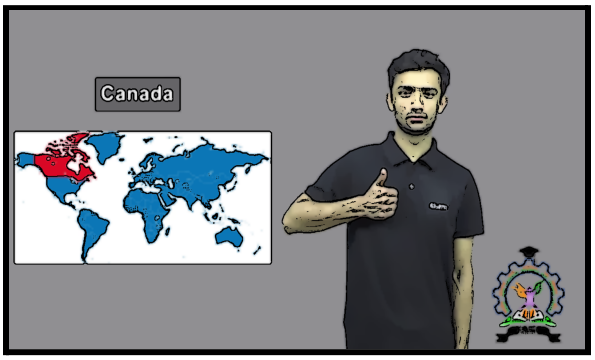

Fig.7 An example of text to Gesture

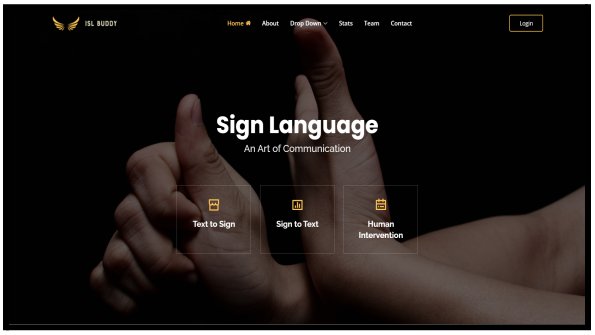

Fig.8 Final Web App

V. LIMITATIONS

As this model gathers the relative coordinates of the movement, therefore it's mandatory for the user to be in the proper frame while capturing the video. Model may predict incorrect text for the similar actions. As of now the model is based on just word level; it won't be able to predict the course of actions that depicts a sentence or won't be able to differentiate between 2 words if performed concurrently. As of now, the database doesn't contain every word used in day to day life, hence the model may predict a wrong text for the given action.

On the other hand for text to gesture, sometimes the user won't find the action related to the entered text but he/she can surely raise a concern so the database could be updated and he/she can find that particular word next time.

VI. FUTURE SCOPE

Accuracy of gesture prediction can be increased by using more advanced models and also by rigorous training of more data as available. As of now the prediction is word level this can be extended to sentence level too. For predicting a text for the given gesture it takes time this can be done in real time by increasing the processing performance and hence this can be used as an intermediary for conversation between the challenged people.

VII. CONCLUSION

Sign Language Communication deals with sign to text conversion and vice versa. Limited resources for ISL results in problems for gesture to text conversion. Our prototype works for word level sign language not for the course of actions that depicts sentences.

Gesture to text conversion deals with a series of action frames. So to deal with the sequence, the LSTM model is used for taking previous pre-processed frames into consideration. Hence minimising the loss of frame containing an important action that may change the result of sign recognition. The model gives a good 92% of accuracy for prediction of words. Text to gesture uses a database which has various words stored for conversion into its corresponding gesture.

VIII. ACKNOWLEDGEMENT

We extend our gratitude towards Prof. Dhiraj Jadhav sir, without whose motivation and guidance, this project would not have been possible. His constant efforts have proved to be invaluable throughout the project. We are also thankful to Prof. Sandeep Shinde Sir -HOD (Department of Computer Engineering) for their valuable guidance.

IX. REFERENCES

1. Dongxu Li, Cristian Rodriguez Opazo, Xin Yu, Hongdong Li. "WORD-LEVEL DEEP SIGN LANGUAGE RECOGNITION FROM VIDEO: A NEW LARGE-SCALE DATASET AND METHODS COMPARISON". *The Australian National University, Australian Centre for Robotic Vision (ACRV)*
2. Rafiqul Zaman Khan,Noor Adnan Ibraheem. "HAND GESTURE RECOGNITION: A LITERATURE REVIEW". *International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012*
3. Kumud Tripathi, Neha Baranwal and G. C. Nandi. "CONTINUOUS INDIAN SIGN LANGUAGE GESTURE RECOGNITION AND SENTENCE FORMATIOn". *Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)*
4. Prof. Radha S. Shirbhate, Mr. Vedant D. Shinde, Ms. Sanam A. Metkari, Ms. Pooja U. Borkar, Ms. Mayuri A. Khandge. "SIGN LANGUAGE RECOGNITION USING MACHINE LEARNING ALGORITHM". *International Research Journal of Engineering and Technology (IRJET)* Mar 2020
5. Munir Oudah , Ali Al-Naji, and Javaan Chahl. "HAND GESTURE RECOGNITION BASED ON COMPUTER VISION: A REVIEW OF TECHNIQUES"
6. Noorkholis Luthfil Hakim, Timothy K. Shih, Sandeli Priyanwada, Kasthuri Arachchi, Wisnu Aditya, Yi-Cheng Chen and Chih-Yang Lin. "DYNAMIC HAND GESTURE RECOGNITION USING 3D CNN AND LSTM WITH FSM CONTEXT-AWARE MODEL". *National Central University, Taoyuan 32001, Taiwan*

7. Sriram S K, Nishant Sinha. “GESTOP: CUSTOMIZABLE GESTURE CONTROL OF COMPUTER SYSTEMS”. *CODS COMAD 2021, January 2–4, 2021, Bangalore, India*
8. https://main.mohfw.gov.in/sites/default/files/51892751619025258383.pdf
9. https://www.youtube.com/playlist?list=PLFjydPMg4Dapq9vcdmGyHs8uJhiqMgUrX
10. https://takelessons.com/blog/asl-for-beginners
11. https://google.github.io/mediapipe/solutions/pose
12. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss