Sign Language Encoder-Decoder

Prof. Naina Kaushik[1], Sahil Singh[2], Tushar Panchmukh[3], Ayaaz Qureshi[4], Suraj Tiwari[5]

[1]*Assistant Professor, Department of Computer Engineering, MCT's Rajiv Gandhi Institute of Technology, Mumbai*

[2,3,4,5]*BE Student, Department of Computer Engineering, MCT's Rajiv Gandhi Institute of Technology, Mumbai*

***Abstract*—Sign Language is a visual language that requires manual manner to communicate what an individual wants to convey. Sign Language consists of its own grammar and lexicon. It is used by dumb and deaf people to portray what they want to say. It acts as a basic mode of communication for people with hearing impairment and speaking impairment, without which communication between them and normal individuals might be difficult.**

We focus on the development of an automated software system to convert speech to sign language, sign language to text so that the deaf people can effectively communicate among themselves as well as with other people. Creating a desktop application that makes use of a computer's webcam to capture an individual marking gestures for American sign language (ASL), and interpret it into corresponding text and speech in real time. To enable the detection of gestures, we are making use of a LSTM Neural Network (RNN). A RNN is highly efficient in tackling computer vision problems and is capable of detecting the desired features with a high degree of accuracy upon sufficient training.

This Web application will not only help deaf people to understand what a person is trying to ask to them instead it will help a normal person as well, to collaborate and to understand a deaf person.

I. INTRODUCTION

In a country with an estimated 1.3 million people with 'hearing impairment'- as stated by the 2011 Census of India, it is still a struggle for the deaf people to communicate and access education due to the presence of only a handful of schools with sign language interpreters that too located in big cities. It, therefore, becomes a major concern to pave a way for the deaf people that would not only minimize the gap in the deaf people to sign language interpreter ratio but also make them independent through a stage that guarantees self-education and learning of sign language. Our project not only help them to learn as well as we can use as a medium between two peoples as mode of communication.

Sign language is largely used by the disabled, and there are few others who understand it, such as relatives.

Natural gestures and formal cues are the two types of sign language [1]. The natural cue is a manual (hand-handed) expression concurred upon by the user (conventionally), perceived to be constrained in a specific bunch (esoteric), and a substitute for words used by a deaf person (as opposed to body language). A formal gesture is a cue that is established deliberately and has the same language structure as the community's spoken language.[2] More than 360 million of world population suffers from hearing and speech impairments [3]. Sign language detection is a project implementation for designing a model in which web camera is utilized for capturing images of hand motions which is done by open cv. After capturing images, labelling of images is required and then pre trained model is used for sign recognition. Thus, an effective path of communication can be developed between deaf and normal audience. Three steps must be completed in real time to solve our problem:

1. Obtaining footage of the user signing is step one (input).
2. Classifying each frame in the video to a sign.

II. LITERATURE SURVEY

2.1 Survey Existing system

Paper 1:

Title: Sign language translator for mobile platforms

Year Published: 2019

Authors: M. Mahesh, A. Jayaprakash and M. Geetha

Tools Used: Database for further and expand detection set. System camera as input and Neural networks as well
Achievements: An app as a translator they take a picture of a sign gesture and later convert it to a meaningful word
Method: The application captures image using device camera process it and determines the corresponding gesture. An initial phase of comparison using histogram matching is done to identify those gestures that are close to test sample and further only those samples are subjected to Oriented Fast and Rotated BRIEF (Binary Robust Independent Element Features) based comparison hence reducing the CPU time.

Paper 2:
Title: Automated Speech to Sign language Conversion using Google API and NLP.
Year Published: 2019
Authors: Ritika Bharti, Sarthak Yadav, Sourav Gupta, and Rajitha Bakthula.
Tools Used: Usage of Google's speech to text API and NLP
Achievements: So, this system is able to convert any speech to text by verifying the text via videos of sign language.
Method: The automated system takes in speech as input using PyAudio library and converts it to a text which undergoes tokenization and further processing using NLP and is matched with the visual sign word library (videos of sign language) to retrieve individual matched videos and concatenate them all to finally display the merged video on the screen. The performance is compared in both offline and online modes to find that online mode has better state-of-the-art accuracy.

Paper 3:
Title: Static Sign Language Recognition Using Deep Learning
Year Published: 2019
Authors: Lean Karlo S. Tolentino, Ronnie O. Serfa Juan, August C. Thio-ac, Maria Abigail B. Pamahoy, Joni Rose R. Forteza and Xavier Jet O. Garcia
Tools Used: web cam, python cv2.cvtColor module, CNN, Keras, Tensor Flow
Achievements: Easier detection by the system with the help of improved skin color recognition using python.
Method: A high quality web camera is used as a sensor to capture images, the system can recognize the hand from the background clearly using python's cv2.cvtColor which improves the skin color recognition. Finally, the image goes through Keras and CNN to get specific results.

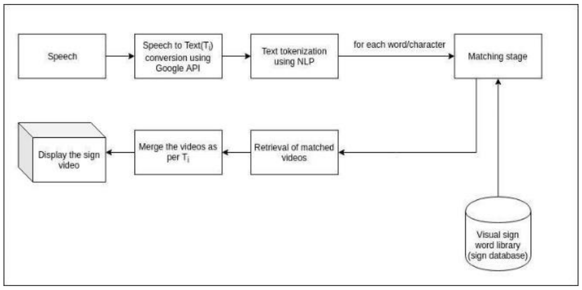
Paper 4:
Title: Conversation of Sign Language to Speech with Human Gestures
Year Published: 2015
Authors: Rajaganapathy S, Aravind. B, Keerthana. B, Sivagami. M
Tools Used: Microsoft Kinect Sensor
Achievements: Being able to recognize the body movement and gestures using Microsoft Kinect
Method: Microsoft Kinect Sensors are used to capture the gesture. It converts the captured video into a skeletal form where the gestures are recognized frame by frame from the skeletal form, once the gesture matches it gives the output.

Paper 5:
Title: Weakly Supervised Training of a Sign Language Recognition System Using Multiple Insance Learning Density Matrices
Year Published: 2011
Authors: Daniel Kelly, John Mc Donald, Member, IEEE and Charles Markham
Tools Used: Videos, Support Vector Machine, Multiple Instance Learning, HMM
Achievements: System is able to extract and recognize gesture patters from videos and images. It is able to give sentences as an output.
Method: System extract sign gestures from a given video source, it recognizes the specific hand postures used for sign language and puts it through a temporal gesture HMM and SVMs to be able to make a clear word or statement to give an output.

III. PROPOSED METHODOLOGY

Sign language encoder-decoder works in two parts:
1) Sign to Text
2) Text to Sign
For the first step sign to text the following methods are being fallowed
1) Image Acquisition

2) *Hand Region Segmentation & Hand Detection and Tracking*
3) *Hand Posture Recognition*
4) *Display as Text & Speech*


Speech
Speech to Text(T_i) conversion using Google API
Text tokenization using NLP
for each word/character
Matching stage
Display the sign video
Merge the videos as per T_i
Retrieval of matched videos
Visual sign word library (sign database)

Flow Chart: The flow chart given below describes the steps used in completion of the objectives of the step.

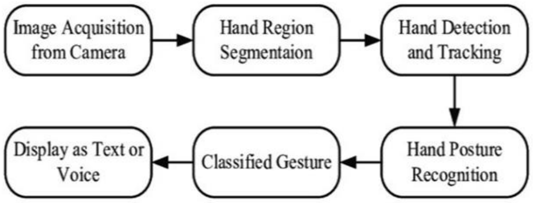


Fig. 1 Workflow of sign to text Proposed Model

1.Image Acquisition The gestures are captured by the help of web camera. This OpenCV video stream is used to record the entire signature duration. The frames are separated from the stream and processed as grayscale images with dimensions of 50*50. This dimension is consistent throughout the project as the whole dataset is sized exactly96+ the same

2.Hand Region Segmentation & Hand Detection and Tracking The captured images are scanned for hand gestures. This is a part of preprocessing before the image get used to the model to obtain the prediction. The segments containing the gesture are highlighted in a better way. This increases the chances of prediction several times.

3.Hand Posture Recognition The pre-processed images are provided to the keras RNN model. An already trained model generates a predicted label. All the motion gesture labels are allocated with a probability. The label with the highest probability is finalized to be the predicted label.

4.Display as Text & Speech The model collects the detected gesture into words. The recognized words are changed appropriately into the corresponding speech using the pyttsx3 library. Speech synthesis output is a simple solution, but it is an invaluable feature because it gives the impression of a real spoken conversation.

For the 2nd Step Text to Sign the fallowing steps are fallowed:

1) *Tokenization of text using NLP*
2) *Matching the visual sign word library*
3) *Merging videos as per T*
4) *i and final display*

Flow Chart: The flow chart given below describes the steps used in completion of the objectives of the step.

Fig. 2 Workflow of Speech/Text to Sign Proposed Model

1.Tokenization of text using NLP Natural Language Processing, popularly known as NLP has brought a revolution in the computation world by providing a means through which computers can understand the human language. This text undergoes tokenization to separate each word individually which may include various forms of the word as per the grammar rules but the root word remains the same such as growing and grown contain the same root word grow but used differently in English. Therefore, the utilization of Stemming and Lemmatization arrives into the picture. Both these techniques are normalization techniques for text or words and are required for early processing of textual matter in order to make it useful for fur- their processes.

2. Matching the visual sign word library For every word/character from the processed text received after the second stage of the application, we perform a matching operation us- ing the tags in the visual sign word library for the videos present in its sign database. Whenever a match is found, the matched video is retrieved from the sign database +and move it to the desired place.

3). Merging videos as per Ti and final display

The individual matches obtained from the last stage are then combined or merged for final display on the application. The operations like merging are here performed using the moviepy library of Python which performs several useful functions for videos

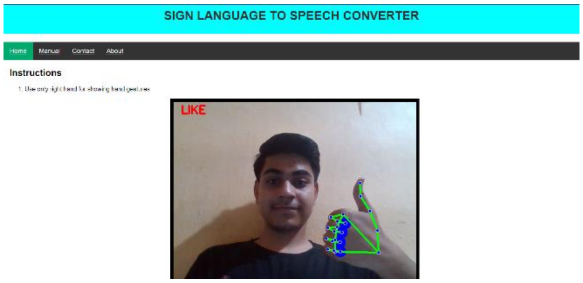
The methodology works on the algorithms area explained as fallows: *RNN for Detection*

RNNs can be used for both sign-to-text and text-to-sign tasks in a sign language translation system. For sign-to-text translation, an RNN can be trained to process the sequence of sign language gestures and generate the corresponding text transcription. As described earlier, this can be done using a sequence-to-sequence RNN architecture with an encoder RNN and a decoder RNN.

For text-to-sign translation, the RNN can be trained to process the input text and generate the corresponding sequence of sign language gestures. This can be done using a reverse sequence-to-sequence RNN architecture, where the input sequence represents the words in the text and the output sequence represents the sequence of sign language gestures. The input sequence is fed into the encoder RNN in reverse order, and the decoder RNN generates the sequence of sign language gestures

IV. RESULTS AND DISCUSSION

Sign Language to Text converter is built using Flask Framework of python and Text/Speech to Sign Language Converter is built using Django Framework of python. In Sign Language to Text Converter coordinates of the hand are tracked using CV2 and MediaPipe module, and matched with the hand gesture available in the database to train and test the model, as shown in figure 1.



SIGN LANGUAGE TO SPEECH CONVERTER

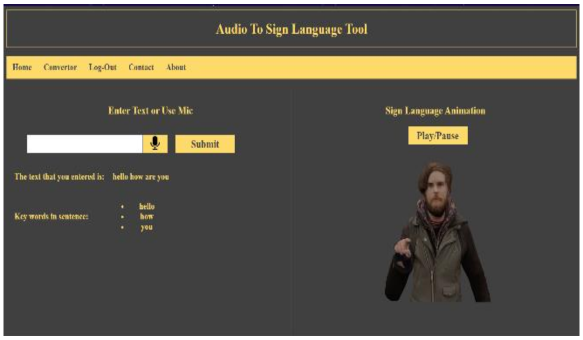
Instructions

1. Use only right hand for showing hand gestures

Fig. 3 Output of Sign Language to Text converter

The model is able to decode the sign language of alphabets and some words to communicate.

In Text/Speech to Sign Language Converter, JavaScript Web Speech API is used for speech recognition, Natural Language Toolkit (NLTK) is used for text processing, and Blender 3D tool is used for creating 3D animation of a character. Various mp4 videos are captured of a character for different actions, after getting input from the user the text is lowered, tokenized, and then converted into tags using NLTK library, after that stop words are removed and lemmatization of words are done, then the animation for every word is played continuously to represent the sign language, as shown in figure 2.



Audio To Sign Language Tool

Enter Text or Use Mic

The text that you entered is: hello how are you

Key words in sentence:

- hello
- how
- you

Sign Language Animation

Fig. 4 Output of Text/Speech to Sign Language Converter

In this model, we can give input in the form of speech as well as text to get sign language.

Lemmatization	
Change	Change
Changing	
Changes	
Changed	
Changer	

Fig. 5 Lemmatization Process

V. CONCLUSION

This Application will definitely help disabled person and there friends, relative, co-workers etc. to get engaged in conversation without need of any means of translator for both of them, by using technologies such as Django, Flask, NLTK, OpenCV, MediaPipe, etc., we were able to create this user friendly application which is faster and has good amount of accuracy.

ACKNOWLEDGMENT

This paper and Research behind this paper would not have been upto the mark without the support of Professor Naina Kaushik. She always kept us motivating and helped us in finding the best resources to carry forward the research paper at its best quality for delivery.

Each and Every Team member who were working on the project had there own expertise and had a different way of solving problems which we were facing while building the project, this helped us to have a lot of different perspective to tackle and solve the Problem Statements.

REFERENCE

[1] 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, pp. 1168-1171, doi: 10.1109/TENCON.2018.8650524.

[2] 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, pp. 1176-1181, doi: 10.1109/ICACCI.2017.8126001

[3] Kelly, J. Mc Donald and C. Markham, "Weakly Supervised Training of a Sign Language Recognition System Using Multiple Instance Learning Density Matrices," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 41, no. 2, pp. 526-541, Ap

[4] Gunasekaran, K., and An, R. Sign language to speech translation system using pic microcontroller.

[5] Rajaganapathy, S., Aravind, B., Keerthana, B., and Sivagami, M. Conver- sation of sign language to speech with human gestures. Procedia Computer Science 50 (2015), 10–15.