American Sign Language Translator Using Machine Learning

Vijay Kumar Sharma
Department of Computer Science and Engineering

Naman Malik
Department of Computer Science and Engineering

Rachit Arora
Department of Computer Science and Engineering

Riddhi Jain
Department of Computer Science and Engineering

Prachi Gupta
Department of Computer Science and Engineering

Abstract— **Communication plays a vital role in interacting with others, so building a program for deaf and dumb would be boon for them in communicating. Unfortunately, there are many disabled people in the world and there is possibility of disable people in the next generation. So, the need of this application would be a great help for the disable person to easily communicate with other people. This paper propose an application, the application is based on speech recognition by using PyAudio library this would provide the output in the visualized format of the spoken words that can be easily understand by the deaf and the use of image processing with use of OpenCV would provide the output in the audio format or text that can be easily understand by the dumb. This program can be used in all communication areas whether it may be public places such as government offices, banks, Railway Station etc. or private areas such as Shops, Schools, Tourist places etc.**

Keywords— Speech Engine, Image Processing, OpenCV, Speech Recognition, Gesture Recognition, American Sign Language, Translator.

I. INTRODUCTION

Sign Language plays a very important role for communication for about 70 million deaf and dumb people as this is their first language. This sign language uses body parts for conveying the meaning to the other person. This language is considered to be the oldest one because this was also used in the stone age period. But as the evolution occurred the language is also structured by the time. This sign language can involve the simultaneously combining shapes of hand, movement and orientation of hands, facial expression and arms. Communication is the biggest problem as we don't know what disabled person is saying to us and vice versa. And if you know Sign language but maybe other people don't know. While Travelling, they can't hear voices of vehicles, announcements etc. so sometime they might help you for the help but you can't because you know how to communicate with them. For developing this program requires the use of NLTK, OpenCV, PIL libraries in python. Different algorithms used to propose this program are Image Processing Algorithm and Speech Recognition Algorithm for creating different phase of this program. In "Hand gesture-based TV control system towards both user and machine friendly gesture applications", Shimada et al. propose a TV control interface using hand gesture recognition [4] which was also done by using hand gestures which was a good inspiration to work on such approach. The add on of this study is carried out forward which is in this paper..

II. LITERATURE REVIEW

Sadaoki Furui in his paper surveyed the majority of themes and advances made in the last 50 years of research so as to provide a better technological perspective and an acknowledgement of the fundamental progress that has also been accomplished in this important area of speech communication. According to him we need to study more on human speech process to make better speech processing [1]. D. R. Reddy worked on the computer speech Recognition by directly analyzing the speech waves he directly sent the speech signals to the computer without filters he applied procedures for segmentation and pitch extraction, prosodic parameter determination, phoneme classification [2].

In "Real-time hand status recognition from RGB-D imagery" used most natural way was used to overcome Kinect TM -based interfaces is the need for persistence in order to interact with virtual objects and they improves inter-face reactivity is to employ a vision module able to recognize simple hand poses with dataset of more than 30 thousand depth-RGB (Red Green Blue) image pairs of hands that is being made publicly available. They also proposed a method with 98% accuracy and is very responsive in the results shown [5].

In "A study on static hand gesture recognition using moments." the author presented a system that detects/finds the hand region through the skin color identification & obtains the binary silhouette (outline or shape). The feature of moment of normalized hand gesture is classified using a minimum distance classifier [6]. In the paper "Digital Art using Hand Gesture Control with IOT" the major concern of the paper was to connect the handicapped with real world with great esteem. In this a new approach was introduced with the help of IOT [7].

In paper "American Sign Language Recognition with the Kinect" they investigated on the potential of the kinetic depth-mapping camera for the recognition of the sign language and verification for educational games for deaf children. They collected a total of one thousand American Sign Language (ASL) phrases across both systems and they got about 76.12% and 51.5% sentence verification rates when all the users are standing and seated respectively on the adult data taken [8]. Helene Brashear created an automatic sign language recognition system for the Copycat

which is America Sign Language (ASL) game. She believes that she can improve the accuracy of this system by characterizing and modelling disfluencies found in the children's signing and worked on that [10].

In the article "Real-Time Hand Gesture Recognition Using Finger Segmentation" the author's method showed better performance than a state-of-art method on another data set of hand gestures. They extracted the hand region from the background with the background subtraction method this resulted in segmented fingers and palm for the recognition [11].

III. Proposed Algorithm

There are many algorithms and programs for the speech recognition and gesture recognition. But they work differently in this we have joined the different approaches to make all the multiple events occur at same place and also this will offer a single package for the communication of both the disable and normal person, where both of them will have a medium to interact with each other.

This has two phases of working-

- Speech\Text to Sign Language
- Sign Language to Text\Speech

A. Speech/Text to Sign Language

There are many algorithms which convert your speech to text and also there are speech engines of Google and Microsoft which also do the same. We have taken help of the existing speech engines and made some advancement in the existing speech engines where with the use of this speech engine we have recognized the sentence of the user who is speaking and then used every word of the sentence to map them with the images associated with them.

We have used file handling for fetching of all the images required because database connectivity can result in the delay of the process and can also increase the time complexity of the program. We have also used the existing media player in the system to eliminate the problem of installing any other specific media player as different platforms have different media player this will also create the flexibility to implement this on any of the system in respective to operating system and system configuration.

We have used speech engine of Google in the replacement of PyAudio because PyAudio is not generating the correct output as desired and also it is consuming more time.

The steps of working of this phase are-

- Speech Recognition.
- Extraction of the sentence.
- Splitting the sentence to get all the words spoken.
- Storing all the text in the array for further processing.
- Linking all the text with their associated images in the data set.
- Fetching all the images.
- After fetching all the images then the images will be processed using OpenCV (python library).
- In the image processing all the images will be converted into a common frame which will be calculated by taking the average of the dimensions of the all images.
- After image processing all the images will be stitched to for a video.
- A video will be generated.

Now generated video can be played in any of the existing media player.

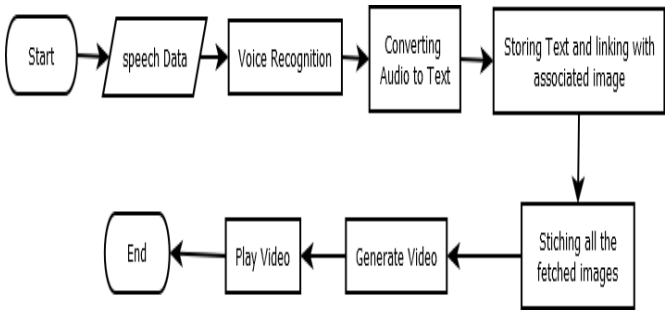


Fig. 1. Diagram of Speech to Gesture Video

B. Sign Language to text/speech.

For the conversion of the sign language we have considered that the user will be using only hands for the communication and eliminated the facial gestures. For the hand gesture recognition, we have used "Palm and Finger Segmentation".

Now on the basis of segmentation all the fingers on the hand image are discovered and recognized. Simple Rule Classifier is used to accomplish the hand gestures. Performance of this method depends on the data set of hand images. Most data sets for automatic sign language recognition are scripted data sets collected in the laboratory by the researchers Subtraction [3,10].

This method is used for detecting the hand region. To recognize the figure the palm and fingers. Then we will use simple rule classifier for the hand gesture classification after recognition of the fingers. SVM and CRF are sophisticated but complicated. In the initial stage the user needs to use data glove for acquiring hand gesture data. In the speech engine we can use SAPI, PyAudio, Google speech to text.

Steps for the hand gesture recognition-

- Hand Detection
- Segmentation of finger and palm.
- Finger recognition.
- Calculation of multiple finger point and angles.
- Recognition of hand gesture.
- Displaying text of that action.
- Now saving that text in a string form and appending if any new word added.
- Sending that string to speech engine.

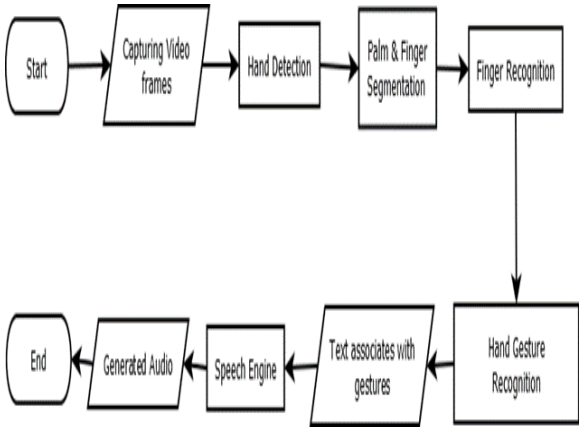
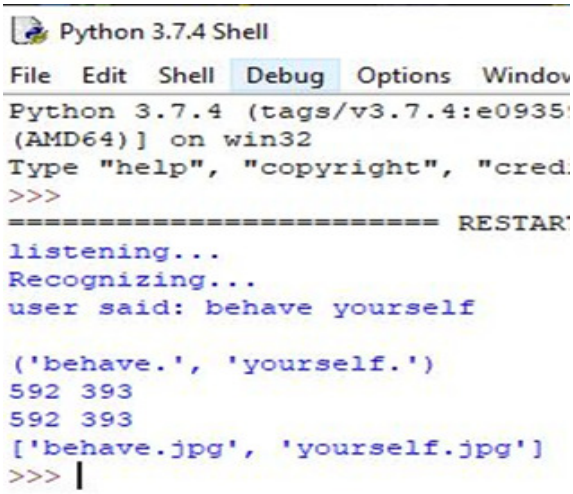
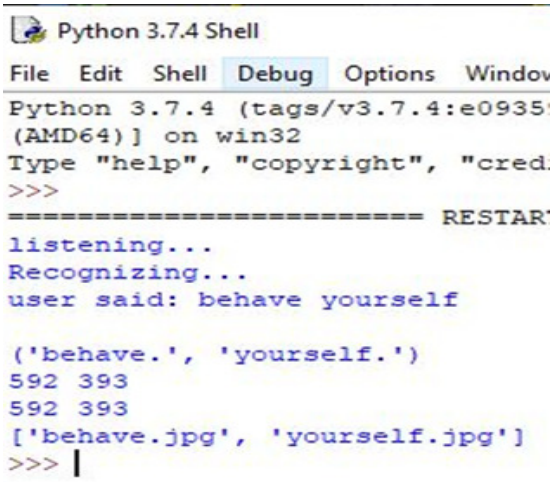


Fig. 2. Flow Diagram of Gesture to Audio

IV. Experimental Result

Proposed technique is implemented and verified on our local system. In our experiment we have created a python program for all the working. This python program is running as expected. We took a small Data Set of gesture images for the testing of speech to gesture part and we have tested lots of sentences and the video is generating as expected also the video is understandable for all the sentence.





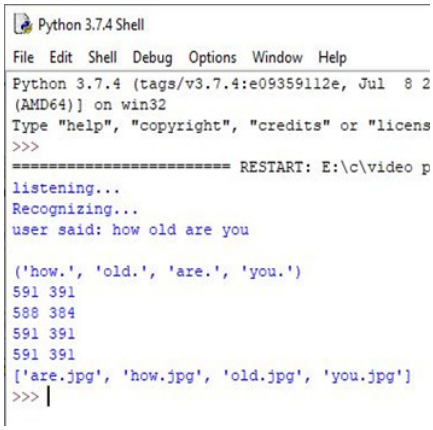


Fig. 3. Testing speech to video

For the gesture to text part, we have created a python program which registers your hand gesture inputs and converts your gestures into text. In this phase we have tested only small gestures like numbers and hello/hi, because this requires to calculate all the angles and defects of the figures so we have calculated only for such small gestures. After implementation this part is also working as expected to the given real time text output of hand gesture. Both the parts Text to Speech and Speech to text is working as proposed in this paper.

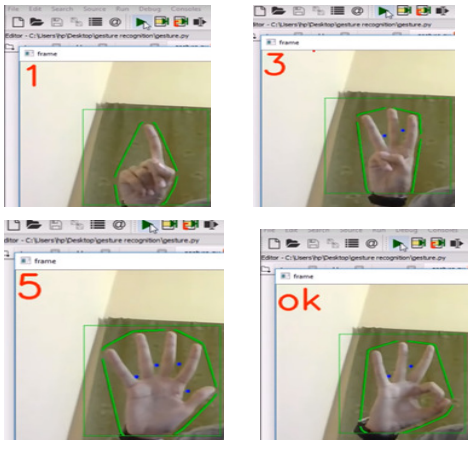


Fig. 4. Testing Gesture to Text

V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a new method of creating an interface for the communication between a disable person and a normal person. This can eliminate the language barrier between the disable and normal person. This approach is applied considering the real time output in the less time. This approach relies on the network for using the speech engines so internet connectivity is required in this approach, offline speech engines of the python has less accuracy and takes more time for pro-cessing, these engines can be used for the devices which have no connectivity options. This proposed method can be implemented in multiple devices irrespective of operating system and system configuration some examples are smart watches, phones and kiosk. According to the experimental results it is clear that by implementing this approach the users will not have any problem for using this program.

This can also be carried out by using any external sensor like lidar sensor or sonar sensor (like google has used in its latest pixel device) for getting more accuracy in the recognition of the gesture. Also for the future this approach can be applied in the our smart watches where a small camera or sensor can detect the movements of the person standing Infront and tell us what the person is saying to us in gestures because smart watches are very compact and usable also they will be much usable if this approach is applied on them.

REFERENCES

[1] Furui, Sadaoki. "50 years of progress in speech and speaker recognition research." ECTI Transactions on Computer and Information Technology (ECTI-CIT) 1, no. 2 (2005): 64-74.

[2] Reddy, D. Raj. "Approach to computer speech recognition by direct analysis of the speech wave." The Journal of the Acoustical Society of America 40, no. 5 (1966): 1273-1273.

[3] Ten Holt, Gineke, and Supervisors Petra Hendriks. "The eye of the beholder: Automatic recognition of dutch sign language." In University of Groningen. 2004.

[4] Shimada, Atsushi, Takayoshi Yamashita, and Rin-ichiro Taniguchi. "Hand gesture-based TV control system − Towards both user-& machine-friendly gesture applications." In The 19th Korea-Japan Joint Workshop on Frontiers of Computer Vision, pp. 121-126. IEEE, 2013.

[5] Bagdanov, Andrew D., Alberto Del Bimbo, Lorenzo Seidenari, and Lorenzo Usai. "Real-time hand status recognition from RGB-D imagery." In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), pp. 2456-2459. IEEE, 2012.

[6] Priyal, S. Padam, and Prabin K. Bora. "A study on static hand gesture recognition using moments." In 2010 International Conference on Signal Processing and Communications (SPCOM), pp. 1-5. IEEE, 2010.

[7] HOD, P. Gnanasundari, K. Jawahar[1], J. Elango, M. Anburaj, and K. Harish. "Digital Art using Hand Gesture Control with IOT."

[8] Zafrulla, Zahoor, Helene Brashear, Thad Starner, Harley Hamilton, and Peter Presti. "American sign language recognition with the kinect." In Proceedings of the 13th international conference on multimodal interfaces, pp. 279-286. 2011.

[9] Athitsos, Vassilis, and Stan Sclaroff. 3D hand pose estimation by finding appearance-based matches in a large database of training views. Boston University Computer Science Department, 2001.

[10] Brashear, Helene. "Improving the efficacy of automated sign language practice tools." ACM SIGACCESS Accessibility and Computing 89 (2007): 11-17.

[11] Chen, Zhi-hua, Jung-Tae Kim, Jianning Liang, Jing Zhang, and Yu-Bo Yuan. "Real-time hand gesture recognition using finger segmentation." The Scientific World Journal 2014 (2014).