International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 04 | April - 2022 Impact Factor: 7.185 ISSN: 2582-3930

REALTIME SIGN LANGUAGE TO AUDIO & TEXT TRANSLATION USING MACHINE LEARNING

[1]Prof.Mayank Mangal, [2]Dishant Deshmukh, [3]Vedant Deshmukh, [4]Smita Vishe

[1]*Head of Department, Department of Information Technology, Alamuri Ratnamala Institute of Engineering and Technology*
[2,3,4]*Student, Department of Information Technology, Alamuri Ratnamala Institute of Engineering and Technology*

ABSTRACT

Sign language is the language which is used by the people with speech & hearing disability. It is a technique where people use our hands, palms, fingers to communicate with each other. Generally, it is difficult for the normal people to understand this language. This project deals with the translation of sign language into audio and text. Steps in recognizing and translating the sign language are described in this study. The proposed model is able to recognize a total 30 different signs included 26 alphabets of ASL(American Sign Language) with an average accuracy of 95% in real-time. Real-time sign language recognition without using any external devices, sensors make it easy and comfortable to use. In the proposed model, we have used only a computer with webcam.

KEY WORDS

Machine Learning, Sign Language, Mediapipe, American Sign Language

INTRODUCTION

A language is nothing but a medium of communication which is generally used for communication, sharing ideas, etc. Sign language is a language designed to make the communication easy for speech and hearing impaired people to communicate with others. Although the recognition of this language is difficult for the normal people which make the specially abled people feel isolated from the normal world. But if we create a in-between medium which can act as a translator then it can be helpful for both.

There are many types of sign languages Indian Sign Language, French Sign Language, British Sign Language, American Sign Language, Indonesian Sign Language, etc. The language used in the proposed ml model is ASL i.e. American Sign Language.

The aim of this project is to build and implement a computer vision model which is capable of recognizing and translating sign into audio and text in real-time so, the blind people can also be benefited with this project. The proposed model can recognize 26 ASL alphabets and 4 user defined signs a total 30 signs.

In this project our basic focus is on creating a model which will recognize hand gestures in order to form and convert them into a complete word by combining every gesture.

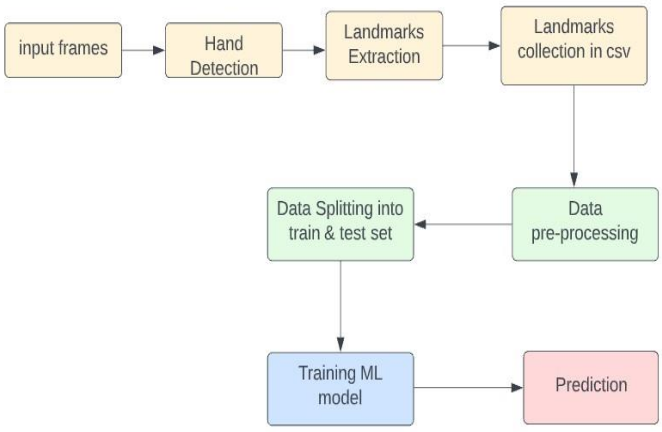
OBJECTIVES

Design and deploy a machine learning model which is capable of predicting sign language in real-time accurately and form a sentence with the predicted words and convert sentence into an audio

RELATED WORK

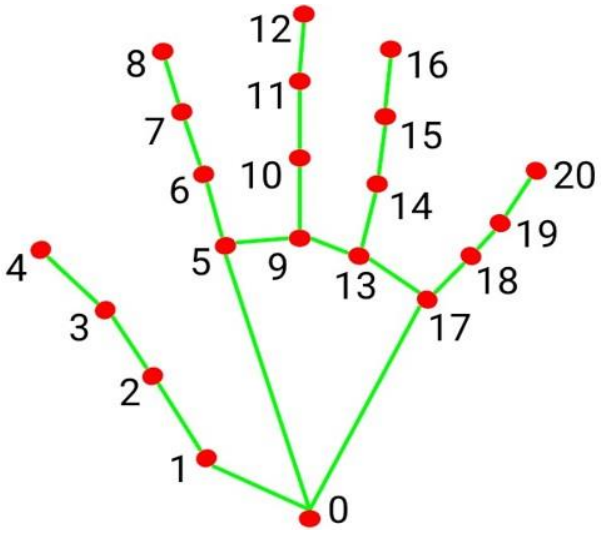
In the field of machine learning the hand gesture recognition is comparatively difficult as it need to track hands. The first research on sign language recognition was published in the year 1991 by Murakami and Taguchi [1] using neural network. After this numerous researches was done in the field of computer vision. Support Vector Machine(SVM) has been used by Tharwat et al.[2] which was showing the better results and Elakkiya et al.[3] used a combination of SVM learning and boosting algorithm to propose a model for subunit recognition of alphabets. The system fails to predict 26 alphabets but they obtained the accuracy of 97.6%. M. Geetha and U. C. Manjusha[4], proposed a model where they use 50 samples of every characters in a vision based recognition of Indian Sign Language characters and numerals using B-Spline approximations. The B-spline curve undergoes a series of smoothening process so features can be extracted. SVM is used to classify the images and the accuracy was 90.00%.
From all the previous research it is clear that to recognize sign language accurately we require a huge training and test dataset and to collect data we used an open-source framework by Google called Mediapipe which is capable of recognize human body part accurately and to train the model we used Random Forest classifier.

ARCHITECTURE



MEDIAPIPE FRAMEWORK

Mediapipe is a Google open-source framework used to track human pose, hands, etc. Mediapipe hands is a reliable hand and fingers tracking solution. It uses Machine Learning technique to understand and marks 21 points on hand in a frame-by-frame approach. This approach is beneficial for the real-time approach as it increases performance. Mediapipe Hands uses an integrated ML of many models working together: The hands detection model which works on the full image and returns the directed hand binding box. Hand gesture model applicable to image-cut region defined by a hands detector once returns 3D hand key points with high reliability.



Multi-hand Landmarks using MediaPipe

The mediapipe hand tracking solution[5] has two ML model in its backend i.e. BlazePalm detector and Hand Landmark Model. The palm detection model detects a palm from the frame and after running palm detection over the frame, hand landmark model performs precise landmark localization of 21 co-ordinates inside the detected hand regions via regression.

- 21 hand landmarks consisting of x-axis, y-axis, and relative depth.
- A hand flag indicating the probability of hand presence in the input image.

PROBLEM STATEMENT

There are many languages in the world and learning those languages can be easy but, sign language is the language which require a lot time to learn although most of the people don't know how sign language works we are creating a model which will help them to recognize sign language.

SCOPE OF WORK

Speech impaired people use fingerspelling and gestures to communicate. Normal people face difficulty in understanding their language. Hence, there is a need of a system which recognizes the different signs, gestures and conveys the information to the normal people. It fulfills the gap between differently abled people and normal people.

METHODOLOGY

There are various steps involved in creating the proposed model and the first step of this is data collection/acquisition. The data collection can have different approaches the two widely used method are:

- Glove based approach
- Vision based approach

In the proposed model we have used vision-based approach. In this technique to detect and track hand we have used Mediapipe. Mediapipe is a google technology generally used for detection of pose, hands, face, etc. In our research we have used mediapipe hands recognizer to collect data and to predict it in real-time.

The first step of the proposed model is to collect data. In our model we have used web camera and collected data in realtime. The web camera is working with mediapipe hands detector which track and make 21 points on hands. Here instead of images we collected the points which was forming on hands along with x-axis & y-axis. We have collected a total of 85,000*21 points on an average 2800*21 points for each character a total of 30 character/sign.

The second step is selecting the machine learning algorithm which gives the best accuracy. In the proposed model we have used random forest algorithm which was giving the accuracy of 98% on test set and about 90-95% on real-time.

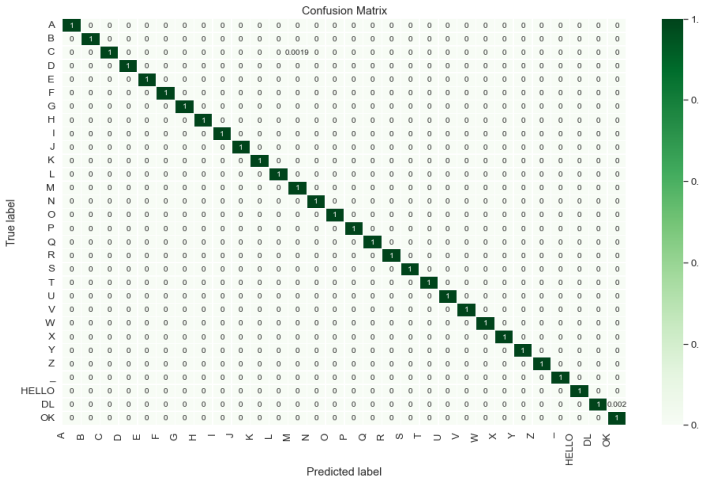
The third and the most important step is to implement the saved model which can run on real-time using the computer vision technique.

The real-time prediction using computer vision is then predicts words and the predicted word is then appended to a string to form a sentence.

The sentence is then passed through the Speech Synthesizer. Speech Synthesizer[6], is the artificial computer-generated voice simulator which simulate human like voice. The sentence is then passed through speech synthesizer which convert the sentence into an audio and the audio is in the computer generated human voice.

RESULT

The average accuracy after applying random forest algorithm is 99.56% and after applying SVM the accuracy is 99.45% on the test set. Both the accuracy is almost same. When the model is tested on real-time the random forest algorithm showing slightly better accuracy of about 95%. The model is now able to predict/recognize character accurately. The predicted characters are then appended to string one by one to form a sentence. The sentence then is converted into audio using speech synthesizer which then give the output in human like computer-generated voice.



CONCLUSION

In the proposed model, Sign Language Recognition using Medaipipe framework, Computer Vision and Speech synthesizer is very successful and the accuracy during testing on real-time is very accurate. The model using Random forest algorithm gives the accuracy more than 99% on test set and average accuracy during real-time testing was about 95%. Our model predicts 21 landmarks with the help of webcam which is then passed through the ML model using computer vision for prediction on real-time and then it gives an output in text & audio format, as it doesn't require any hardware or sensors so it can be easily deployed to various devices.

FUTURE WORK

Our present model can identify 30 signs i.e. A-Z, Hello, Ok, & Delete(to delete wrong prediction). We look forward to add more signs in our dataset so it can predict more words.

REFERENCES

[1] Murakami K, Taguchi H. 1991. Gesture recognition using recurrent neural networks. In: Proceedings of the ACM SIGCHI conference on Human factors in computing systems, pp 237–242.

[2] Tharwat A, Gaber T, Hassanien AE, Shahin MK, Refaat B. 2015. Sift-based arabic sign language recognition system. In: Springer Afro-European conference for industrial advancement, pp 359–370.

[3] Elakkiya R, Selvamani K, Velumadhava Rao R, Kannan A. 2012. Fuzzy hand gesture recognition based human computer interface intelligent system. UACEE Int J Adv Comput Netw Secur 2(1):29–33 (ISSN 2250–3757).

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

[5] Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C. L., & Grundmann, M. 2020. MediaPipe Hands: On-device Real-time Hand Tracking. arXiv— preprint arXiv:2006.10214.

[6] https://en.wikipedia.org/wiki/Speech_synthesis