International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 03 | March - 2021 ISSN: 2582-3930

Sign language to speech and text conversion with augmented reality

Priyadharshini. M[1], Sivashree. V[2], Gayathri. J[3]
Final Year Students[1][2],Assistant Professor[3],
Department of Electronics & Communication Engineering,
Prince Shri Venkateshwara Padmavathy Engineering College ,Chennai 127

Abstract

Communication provides interaction among the people to exchange the feelings and ideas. The deaf and dumb community suffers a lot to interact with the community. Sign language is the way through which the people communicate with each other. In order to provide interaction with normal people there is a system which can convert the sign languages to the understandable form. The purpose of this work is to provide a real-time system which can convert Sign Language (ISL) to the text. Most of the work based on hand crafted feature. The deep learning approach which can classify the sign using the convolutional neural network is introduced. First, a classifier model using the signs using the Kera's implementation of convolutional neural network using python is made. Then another real-time system which used skin segmentation to find the Region of Interest in the frame which shows the bounding box is processed. The segmented region is feed to the classifier model to predict the sign. The classifier is found to be improving with different background and the angle of the image captured.

Keywords: Sign language Recognition, Skin Segmentation, Kera's, TensorFlow, Convolutional neural network, Real time Recognition.

1. Introduction

In the world, different languages are used among the people to provide communication while talking about physically impaired people both deaf and dump community also use different sign languages. The different languages are American Sign Language, Chinese sign language, Indian sign language etc. A real time system among these community will allow the communication barrier among them. Once it has been converted to using the Computer Vision approach then it can be converted to any language.

There are many kinds of research undergone by in this are to build an efficient and accurate system. The previous works done by the researchers using the hand crafted feature but having limitation and used special conditions. The most works are based on pattern recognition, feature extraction based on HOG, SIFT, LBP, etc. But the system using a single feature is not sufficient in most of the cases and the Hybrid approach are introduced to solve this problem. But for a real-time system we need faster methods to solve our problems. Now a day our computers are improved with the speed of processing using parallel implementation. In most of the time our system utilizes a single core for solving a problem. Using the GPU system, the problems can be solved by parallel computing and the number of cores is higher than the CPU system. Using the Deep Learning approach, we can model a self-learning system for our needs. Convolutional neural network is one of the trending deep learning systems which is capable of solving any kind of computer vision problem. In our method we used a Region of interest-Convolutional neural network for the real time Sign Recognition using skin segmentation.

2. Related works

This project paper [1] was mainly aimed to produce a prototype for checking the feasibility of recognizing gestures using image processing. The designed project converts hand gestures into speech which can be understood easily by normal people. The system tells the meaning through the audio output. Programming logic used in the proposed system will not be able to identify more gestures. Multilingual speech enabling system can be developed according to the paper. Programming methods used in the system can be enhanced.

The technical aspect of the recently started ViSiCAST aims to provide deaf citizens with improved access to services and facilitates through sign language. At the core of the project [2] are enabling technology work package, concerned with the further development of the language technology and Avatar technology established earlier projects. These activities support work packages covering three different application areas: multimedia and WWW; face-to-face transactions; and broadcasting. It gives an overview of the main application areas, before considering more detail to areas of supporting technology development: language and notation, and virtual human signing. High end processors are needed for this system and thus implementation is tedious.

Enhancing the human interactions with the digital world and evaluating the performance change that occurs with different modalities. The virtual keyboard specified in this paper [3] is based on menu selection with seven commands that allows to type 30 different characters and allows correcting the errors using clear button. The performance of the proposed system is evaluated by Information Transfer Rate (ITR) at both command and application levels. The performance rate also depends on the user's experience. The measured performance rate of the system is found to be less.

In this paper [6], the main process carried out is the head detection and eye tracking. This paper provides various methods and approaches for the tracking and detection. It helps in interaction between human and computer. The paper [9] is based on the virtual reality and this methods are used in television programs.

This paper [14], represents the Indian sign language recognition for alphabets and its respective voice transformation. This paper includes HMM and hardware like raspberry pi for sign recognition. The paper [13] describes the sign language recognition using artificial neural networks. In this paper [15], the raspberry pi is used in the image processing for 26 alphabets and 9 digits.

3. Proposed method

The sign language is a complex system which includes the background noises. The efficient way is to perform convolutional neural network on the image to increase the efficiency of classification and for real life application.

Most of the object detection problems train the model using the image data set along with a bounding box mapping. The marking of bounding box for each image is costly. In addition, with that we proposed a region of interest predictor using the skin segmentation. From the segmented bounded region, we crop the image and feed to the classifier for prediction.

3.1 Block diagram

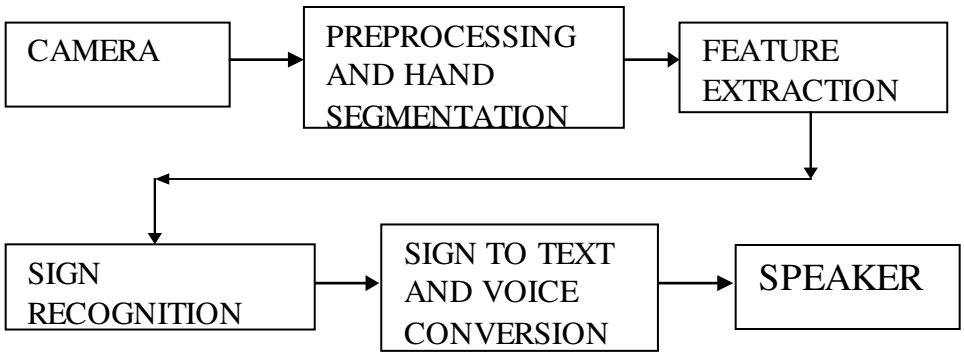

Figure 3.1 Block diagram of proposed system

3.2 Working

The signs shown by the dumb people are processed with the help of digital image processing. The captured image from the camera is subjected to preprocessing to remove the noise. The image is separated from hand colored pixels and thus mask is generated. The mask undergoes sign recognition through keras model which has a number of images in the database under single classification. The mask by matching with the database is labeled to the respective sign; text is shown in the live image streaming window. The speech output is heard through the speakers.

3.3 Implementation

Implementation of proposed system needs algorithm to process the image and to recognize the gestures, using Python and OpenCV.

The program implemented has the following procedure

- Captures the gestures through a cameraϖ Sets-up an interface to convert from RGB to HSV colour space andϖ determine the HSV values for skin colour
- Uses the HSV values to reduce the image to black and white
- Produces a gray scale image in preparation for labelling and blob detection
- Produces a blob of the hand in preparation for feature extraction
- Extracts features from the blob for gesture recognition
- Displays the text output and produces speech output

3.4 Flow chart

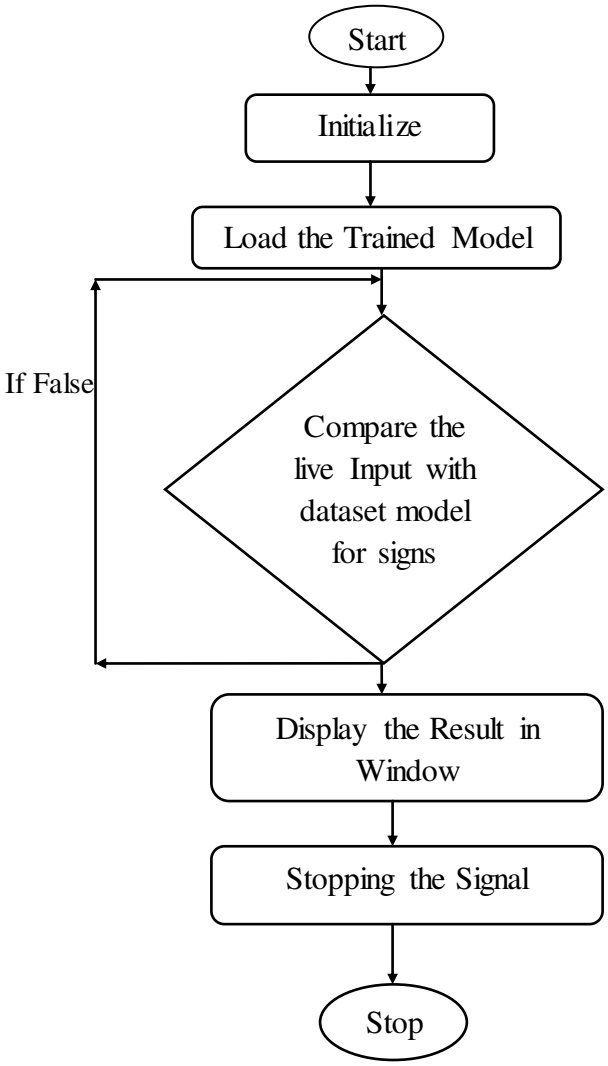

Figure 3.2 Flow chart of proposed system

4. Results and discussion

The image is captured from the camera. The mask window displays the mask of the captured image. The gestures are classified using Python programming and the features are identified. The respective text for the matched sign feature is recognized is and the text is displayed in the screen. Then speech output is heard through speakers.

4.1 Sample output

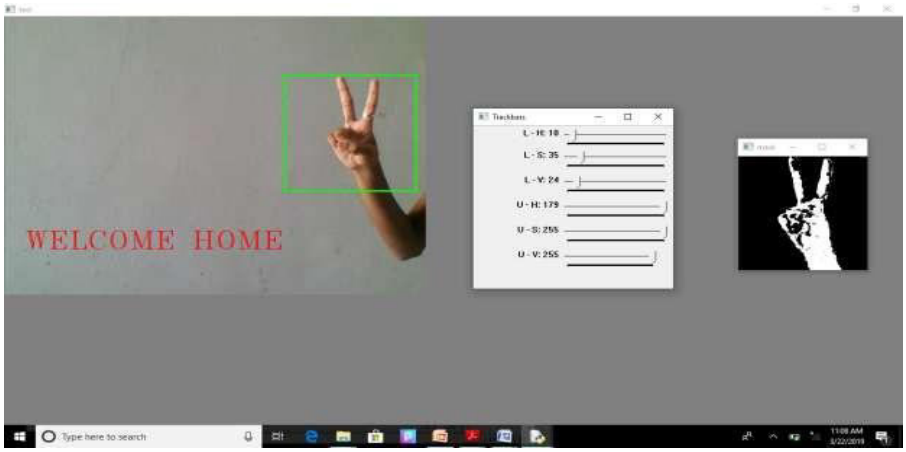

Figure 4.1 The output for the hand gesture (WELCOME HOME)

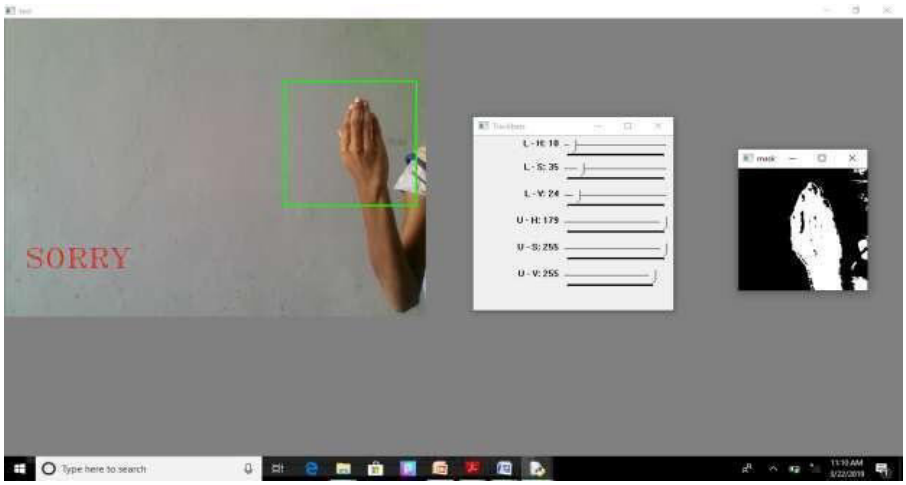

Figure 4.2 Screenshot for hand gesture (SORRY)

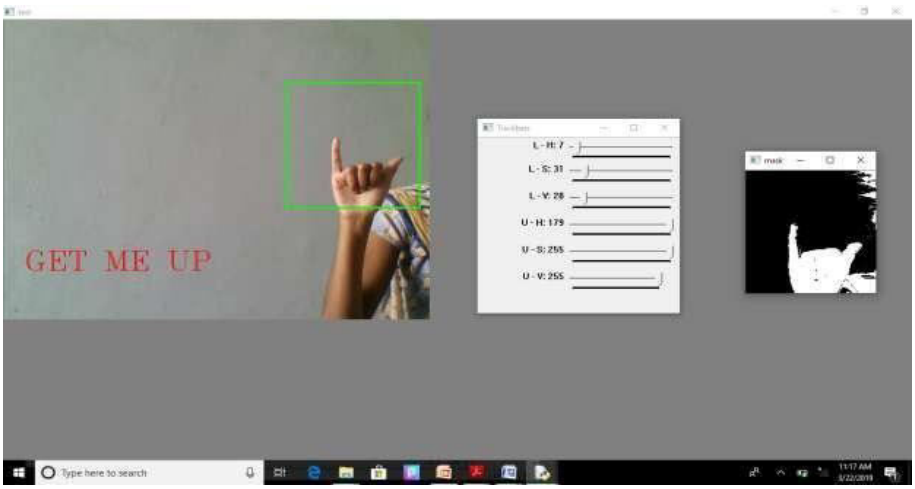

Figure 4.3 Screenshot for hand gesture (GET ME UP)

5. Conclusion

The effective use of Augmented Reality technique enables the patient to communicate with the normal people. The developed system not only benefits the dumb people but also patients who are unable to speak due to various reasons such as aging paralysis. The system reduces the number of hardware components used for communication. Real time gestures given by the dumb people are processed. The system is portable anywhere and easy to access. Thus the proposed system was developed using Python and OpenCV software.

References

1. "Real Time Sign Language Recognition using Python", Supriya Pawar, Sainath Muluk, Sourabh Koli, International Journal of Innovative Research in Computer and Communication Engineering, March 2018.

2. "Virtual Signing: Capture Animation, Storage and Transmission – An Overview of the ViSiCAST Project", J.A. Bangherm, S.J. Cox, R.Elliott, J.R.W. Glauert, I. Marshell, Rankov, M. Wells, IEEE Seminar on "Speech and language processing for disabled and elderly people", London,April 2000.

3. "Augmented Reality Virtual Keyboard including Text to Speech Conversion" T. Maheshwari, S. menagapriya, P. Geetha, Asian Journal of Applied Science and Technology(AJAST), April 2017.

4. Hubert ceccoti "A Multimodal gaze-controlled virtual keyboard" in IEEE transaction on human-machine system, 2016.

5. Hafiz adnan habib, mufti "Real time mono vision gesture based virtual keyboard system", IEEE transactions on consumer electronics, 2006.

6. Amer Al-rahayfeh "Eye tracking and head detection movment" IEE journal of transaction on engineering in health and medicine, 2013.

7. Eda akman aydin, Omer Faruk bay "Region based brain computer interface for a home control application", IEEE journal on communication, 2012.
8. "Vocal Platform for Telephone Voice Portals and Internet Based Interfaces", m-ICTE Badajoz 2003, V. Tomico, N. Morales, E. Campos, J. Tejedor, D. Bolanos, S. Jimenez, J. Garridoy, J. Colas.

9. "Signing for the deaf using virtual humans", J.A. Bhangam, S.J. Cox, M. Lincoln, I. Marshall, M. Tutt, IEEE Seminar on "Speech and language processing for disabled and elderly people", London April 2000.

10. "The HTK Book Version 3.2, S. Young, G. Evergreen, T.Hain, D.Kershaw, G.Moore, J.Odell, D. Ollason, D.Povey, V.Valtchev, P. Woodland, Cambridge University Engineering Department, December, 2002.

11. Seo, B. K. Choi, J. Han, J. H. Park "One handed interaction with augmented virtual objects on mobile device in Proceedings of the 7th ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Application in Industry", 2008.

12. M. Higuchi, T.Komuro, AR typing interface for mobile devices Vision-based hand interaction in augmented reality environment, In proceeding of the 12th International Conference on Mobile and Ubiquitious Multimedia, 2013.

13. V. G Adithya, P. R. Vinod and Usha Gopalakrishnan, "Artificial Neural Network Based Method for Indian Sign Language Recognition", IEEE Conference on Information and Communication Technologies (ICT), pp. 1080-1085, 2013.

14. P Vijayalakshmi, M Aarthi, "Sign language to speech conversion," International Conference on Recent Trends in Information Technology (ICRTT), 2016.

15. Pooja Balu Sonawane, Anita Nikalje, "Text to sign language conversion by using python and database of images and videos" International Journal for Engineering Research in Electronics and Communication Engineering (IJERECE), Vol 5, Issue 2, 2018.