



Sign Language Recognition And Speech Conversion Using Raspberrypi

Ramasuri Appalanaidu CH¹, Nambolu Sai Ramya², Killada Sumanjali³

K Venkata Lakshmi⁴, Kinthali Gayatri⁵

^{2,3,4,5} Student, B.Tech (Information Technology)

¹ Assistant Professor, Information Technology,

Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India

Abstract: *Inability to speak is considered to be true disability. People with this disability use different modes to communicate with others, there are many number of methods available for their communication one such common method of communication is sign language. Sign language allows people to communicate with human body language, each word has a set of human actions representing a particular expression. The motive of the paper is to convert the human sign language to Voice with human gesture understanding. This is achieved with the help of Raspberry pi web camera and speaker. There are a few systems available for sign language to speech conversion but none of them provide portable user interface. For consideration if a person who has a disability to speak can stand and perform in front of the system and the system converts the human gestures as speech and plays it loud so that the person could actually communicate to a mass crowd gathering. Also the system helps visually and speech impaired people to communicate with each other.*

Keywords: *Sign Language, Gesture Recognition, Image Processing, Visually and speech impaired, Voice output.*

I. INTRODUCTION

Sign language is a system of communication using visual gestures and signs, as used by deaf and dumb people. There are various categories in the sign language like ISL (Indian Sign Language), ASL (American Sign Language), BSL (British Sign Language) and etc... But none of the sign languages are universal or international. A person should know the sign language to understand those people, this becomes complicated when a person who has inability to speak or hear wants to convey something to a person or group of persons, since most of them are not familiar with the sign language. Humans however migrate towards technology advancements always expect flexibility in the way they use their system and machinery. At present lots of techniques and modulations are being introduced and are under research to minimize or simplify the complexity in sign language to speech. The paper is being proposed in the aim of minimizing all those complexions and to attain maximum accuracy in conversion of sign language to speech with gestures. Human gestures are an important sign of human communication and an attribute of human actions informally known as the body language. A lot of methods are being in use to track human gestures. To get maximum accuracy and to bring out the system unique a lot of

methods are attempted and best case is user defined actions (gestures) to control the system. For example consider a person who has the disability to speak wants to say "Hello" to a group of people who doesn't know sign language. The user stands in front of the system and waves the hands and system throws out the speech "HELLO"

A. Related Work:

Several different models has been designed and implemented for currency recognition by different authors. In [2], the authors have implemented the system using convolutional neural networks, one of the concepts used in deep learning. They have prepared the dataset for the gestures in American Sign Language.

Vaibhav Mehra in [3], proposed a sign language recognition system for visually impaired people using ORB algorithm. They have implemented the algorithm on American gestures with the accuracy rate of 96% and the runtime of 0.682 seconds. The proposed system is deployed

into the mobile device through which the user can scan the gestures and the output is represented in the form of voice using the mobile speaker.

Albert Mayan [4] proposed a system making using of SIFT algorithm on Android platform. They too deployed the system in Android based mobile phones. The drawback with this is that the SIFT algorithm can be used for features extraction but will not detect the text features. So in the proposed system we are using SIFT and along with that OCR (optical Character Recognition) is used for detecting the text features.

In [5], Mansi Gupta represented a paper of review on gesture detection technique in which they have represented the various techniques that are implemented till date. In [8] they have represented the gesture detection system for

American gestures using artificial vision. They have classified the gestures based on color and text features using the RGB space and Local binary patterns.

PROPOSED METHODS:

The below mentioned are the proposed systems that came from the drawbacks of the existing system.

One of the main disadvantage of the existing system is the Portability. Earlier people used to sit in front of the system and used to make sign languages. Now, in our hardware device, people can make sign language where ever they need and whenever they wanted to convey their message to the opponent. That means, we are making the system portable. As already discussed in the existing system, that accuracy plays a major role in any of the systems. Here, in our hardware, we are using Yolo as a software which increases accuracy of the video captured.

Here we are additionally using a Web camera that captures the gestures done with the hands by the people. In this project, we are using Convolution neural network (CNN). Convolutional neural networks have been one of the most influential innovations in the field of computer vision. CNN is a deep, feed-forward artificial neural network, deep learning models. Voice is generated through the software installed "pyttsx". Programming is done with Python Programming Language.

A. image recognition

Image recognition is done using convolution neural networks(CNN)

Let us consider the use of CNN for image classification in more detail. The main task of image classification is acceptance of the input image and the following definition of its class. This is a skill that people learn from their birth and are able to easily determine that the image in the picture is an elephant. But the computer sees the pictures quite differently.

To solve this problem the computer looks for the characteristics of the base level. In human understanding such characteristics are for example the trunk or large ears. For the computer, these characteristics are boundaries or curvatures. And then through the groups of convolutional layers the computer constructs more abstract concepts.

B. Voice conversion

Speech conversion is done using python text to speech conversion module(pytttx).

It is cross platform text to speech library and platform interdependent. This works offline and does not save voice file in your system. It is mainly useful for people who does not need storing voice files.

The figure below shows the block diagram of proposed system

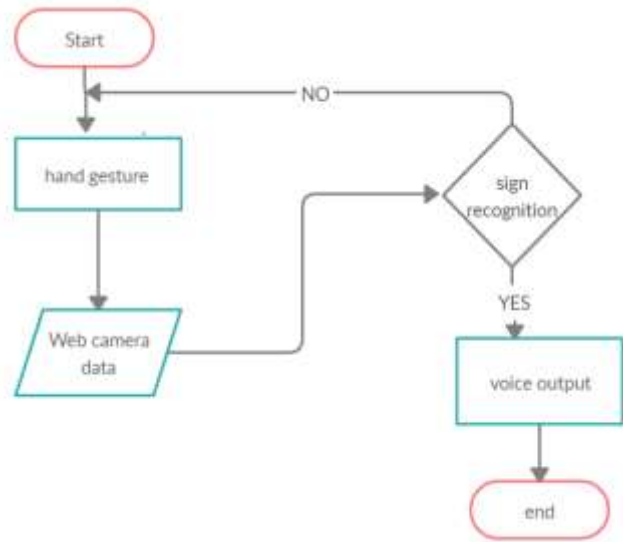


Fig1: block diagram

II. SYSTEM DESIGN

4.1 Hardware implementation of proposed system

A. Raspberry pi

Raspberry pi small and functions like a tiny computer. It has many versions. The version we used is raspberry pi 4, because of its better Computational capabilities , additional ports(usb and hdmi) and more ram space. It has 1.5 GHz 64 bit quad core ARM Cortex-A72 Processor. The OS used is raspbian and the code is written in it such that it reads data streamed from web cam and voice is sent out after execution of code.

The raspberry Pi is a tiny fully functional computer with low cost package. It is provided in various versions. In the proposed system the raspberry PI 3 model is used for implementation. It has a CPU of Quad core 64 bit ARM cortex. It has an internal memory of 1GB and 4 USB ports. Apart from that it has an inbuilt Bluetooth and WiFi. The application is deployed into this tiny computer which is attached to the camera. When a currency note is scanned using the camera, the application in the system will detect the note and provide the results in the form of voice through the speaker.



B. Camera



The Logi-Tech web camera is the one that is deployed on the top of the portable device and is connected to the raspberry PI system. In the proposed system the camera that has a resolution of 16 mega-pixels with USB and night vision is deployed. The camera can scan the gestures during the night time and cost of it is even negligible. The scanned images are send to the raspberry PI and voice is generated out through speakers.

C. Speaker



The speaker is connected to the raspberry PI, which will display the output in the form of voice. The speaker used in the proposed system is a basic model which is used only for audio purpose.

Software Used:

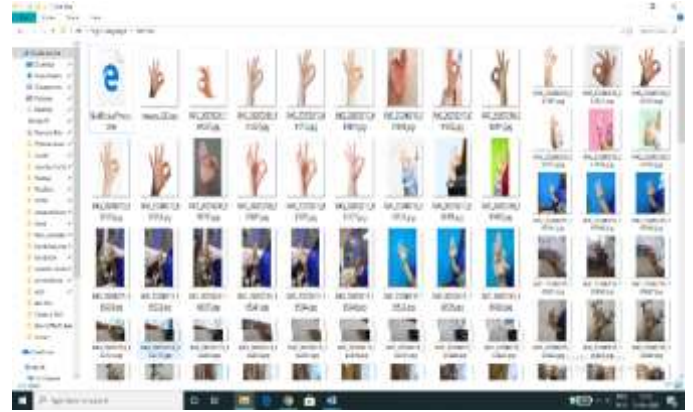
Software is a group of programs that instructs the system to do some specific task as per the commands provided. These programs are built by the programmers for interacting with the system and its hardware. The software required for the proposed system is:

Operating System : Raspberry pi
Scripting Language : Python 3.6.2

. RESULTS

The procedure that is implemented and the set of images that are used in training dataset and the results are represented in this section.

A. Experimental Procedure



The proposed system is deployed on to the device which is attached to the portable device. The camera is mounted on the top of the device and it doesn't rely on capturing the image at a specific degree. In this system the user has to bring the hands in front of the camera and the image will be captured. The proposed system is constructed from the libraries and modules of OpenCV. They are very much efficient and have very good accuracy in getting the results faster.

As represented, they contain the important portions that have unique features that are required to train and predict the gestures. Moreover, in the dataset, we store the important portion of the gestures rather than the entire gesture as they may reduce the efficiency and accuracy of predicting and also reduce the speed of predicting.

B. Visual Results

In this section we have represented specification of different camera devices on which the proposed system is tested and also step by step visual effects of each processing stage. The proposed system is tested on various cameras from VGA which has a pixel resolution of 640 X 480, followed with high resolution pixel.

In the proposed system we have implemented a camera which has an image resolution of 16MP with USB interface and night vision.

The visual process of the system, which represents the test results of all the gestures respectively.

CONCLUSION

In this paper, a sign language recognition system is been proposed for the blind and deaf and visually impaired using CNN algorithm.

In addition, firstly the hand is brought in front of camera and then the hand gestures are made in front of the camera. The Camera recognizes the gestures and then the voice is sent as an output through the speakers. The evaluation results show that the proposed system has a very good accuracy rate with good processing time. However, it has a limitation of differentiating the fake gestures, after acquiring the results through complete analysis by considering different parameters or dimensions in the project. In the future work we will be trying to deploy the techniques related to determining the counterfeit gestures and then display the results.

REFERENCES

- [1] Gupta, Dhiraj-Design and development of a low cost Electronic Hand Glove for deaf and blind, 2nd International conference on computing for sustainable Global Development(INDIA.Com), pp501-505, 11-13 March 2015.
- [2] Sidek, O Hadi, M.A., --Wireless gesture recognition system using MEMS accelerometer, International Symposium on Technology Management and Emerging Technologies (ISTMET), pp 444-447, 2014.
- [3] Mr. Kunal A.Wankhadel, Prof Gauri N. Zade 2, Sign Language Recognition For Deaf and Dumb people using ANFIS, (2014)1206-1209.
- [4] Shape-Based Hand Recognition by Erdem Yoruk, Ender Konuko glu, Bullent Senior Member, IEEE, and Jerome Darbon.
- [5] Manikandan, K., Patidar, A., Walia, P. and Roy, A.B., 2018. Hand Gesture Detection and Conversion to Speech and Text,. arXiv preprint arXiv:1811.11997.
- [6] Padmanabhan, V. and Sornalatha, M., 2014. Hand gesture recognition and voice conversion and dumb people. International Journal of Scientific & Engineering Research, 5((5), p.427.
- [7] Potdar, P.R. and Yadav, D.D., 2014. Innovative Approach for Gesture to Voice Connversion. International journal of innovative research and development, 3(6), pp.459-462.
- [8] Rajaganapathy, S., Aravind, B., Keethana, B. and Sivagami, M.,2015. Conversation of Sign Language to Speech with Human Gestures. Procedia Computer Science, 50,pp.10-15.

