"CONVERSION TO SPEECH OR TEXT FROM SIGN LANGUAGE"

Ms. Supriya N. Mishra[1], Ms. Surbhi K. Nahar[2], Ms. Dhara L. Nayak[3]

Project Guide:- Mr. Vishal Katekar[4]

[1-4]Department of Electronics and Telecommunication, Dr. D. Y. Patil School of Engineering and Technology, Lohegaon, Pune- 412105

Abstract

Impaired persons inability to speak is considered to be true disability. People with this impaired use different modes to communicate with others, there are n 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 or through their gestures 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 and motion capture. This is achieved with the help of CNN (Convolutional Neural Network) Algorithm a motion capture device from Microsoft. There are a few systems available for sign language to speech conversion but none of them provide natural user interface.

For consideration if a person who has a disability to speak can stand perform 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 is planned in bringing high efficiency for the users for improved communication.

Keywords- *Sign Language; Natural User Interface; Gesture Recognition; Rasberry pi; CNN; Image Processing; Python; Pycharm*.

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). But none of the sign languages are universal or international. A person should know the sign language to understand the language; 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.

Recognition of sign language is one of the major concerns for dump and deaf people. Sign language recognition is a research area involving pattern recognition, computer vision, natural language processing. Sign language recognition is a comprehensive problem because of the complexity of the visual analysis of Hand Gesture and the highly structured Nature of sign language. As well as it is considered as a very important function in many practical communication applications, such As sign language understanding, entertainment, and Human Computer Interaction (HCI).

Indian sign language (ISL) is used by the deaf community in India. It consists of both word level gestures and fingerspelling. Fingerspelling is used to form words with letter by letter coding. Letter by letter signing can be used to express words for which no signs exist, the words for which the signer does not know the gestures or to emphasis or clarify a particular word. So, the recognition of fingerspelling has key importance in sign language recognition.

The fingerspelling in Indian sign language consists of both static as well as dynamic gestures which are formed by the two hands with arbitrarily complicated shapes. This paper presents a method for the automatic recognition of static gestures in Indian sign language alphabet and numerals. The signs considered for recognition include 26 letters of the English alphabet and the numerals from 0-9.

Usually deaf people seek the help of sign language interpreters for translating their thoughts to normal people and vice versa. But this system is very costly and does not work throughout the life period of a deaf person. So, a system that automatically recognizes the sign language gestures is necessary. Such a system can minimize the gap between deaf people and normal people in the society. There are various sign languages across the world. The sign language used at a

particular place depends on the culture and spoken language at that place.

II. RELATED WORK:-

Research in the sign language system has two well-known approaches are Image processing and Data glove .The image processing technique using the camera to capture image/video. Analysis the data with static images and recognize the image using algorithms and produce sentences in the display, vision based sign language recognition system mainly follows the algorithms are Hidden Markov Mode (HMM) , Artificial Neural Networks (ANN) and Sum of Absolute Difference (SAD) Algorithm use to extract the image and eliminate the unwanted background noise.

The main drawback of vision based sign language recognition system image acquisition process has many environmental apprehensions such as the place of the camera, background condition and lightning sensitivity. Camera place to focus the spot that capture maximum achievable hand movements, higher resolution camera take up more computation time and occupy more memory space. User always need camera forever and cannot implement in public place. Another research approach is a sign language recognition system using a data glove user need to wear glove consist of flex sensor and motion tracker. Data are directly obtained from each sensor depends upon finger flexures and computer analysis sensor data with static data to produce sentences. It's using neural network to improve the performance of the system.

Its portable device and cost of the device also low. Another approach using a portable Accelerometer (ACC) and Surface Electro Mygram (SEMG) sensors used to measure the hand gesture. ACC used to capture movement information of hand and Arms. EMG sensor placed, it generates different sign gesture. Sensor output signals are fed to the computer process recognize the hand gesture and produce speech/text. But none of the above methods provide users with natural interaction. This proposed system will be capable of performing the conversation without any wearable device instead using the human motion and gesture recognition.

III. DATA SETS :-

- **SIGN LANGUAGES**

1.) VERY GOOD
2.) A
3.) ALL THE BEST
4.) Z
5.) P

Fig(1.1)

Fig(1.2)

Fig(1.3)

Fig(1.4)

Fig(1.5)

IV. PROPOSED WORK:-

- **SYSTEM ARCHITECTURE**

The proposed work is producing speech/ voice to sign language with simple human gestures and motion sensing technology with Image acquisition is the process of capturing the hand gesture images representing different signs. In this step, the image database is created for training and testing the system. The image dataset of Indian sign language alphabet and Sign language recognition systems are broadly classified into two categories: hardware-based systems and vision-based systems. Hardware based systems require the user to wear some devices to extract features describing the hand sign. Cyber glove is a device which is used for extracting the features such as orientation, movements and colour, of the hands. It is widely used for sign language recognition. Vision based systems use digital image processing techniques to extract features and recognize sign. The method proposed in this paper is a vision-based approach in which the user does not have to wear any special device.

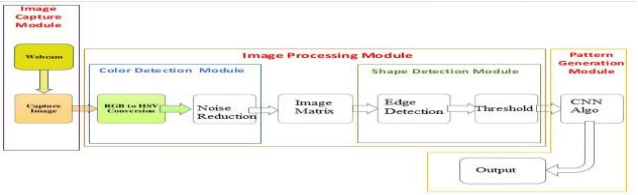

Fig 2.1:- System Architecture

Hand segmentation is the process of extracting the hand sign from the captured image. Efficient hand segmentation has a key role in sign language recognition task. The hand region can be extracted from the background using skin colour-based segmentation. Colour based segmentation is computationally simple and the colour descriptor of an object is invariant to geometric transformations such as translation rotation and scaling. So, colour is widely used as a powerful descriptor for object detection. Colour models represent a colour in a standard way. Different colour models and colour-based systems have been used for skin detection applications. The proposed method for hand detection is applied in the YCbCr colour space. In order to detect the skin colour in the input image it is first converted to YCbCr colour space. YCbCr separates RGB into luminance and chrominance components where Y is the luminance component and Cb, Cr are the chrominance components. RGB values can be transformed to YCbCr colour space using the following equations

$$Y = 0.299R + 0.587G + 0.114B,$$
$$Cr = 128 + 0.5R - 0.418G - 0.081B,$$
$$Cb = 128 - 0.168R - 0.331G + 0.5B.$$

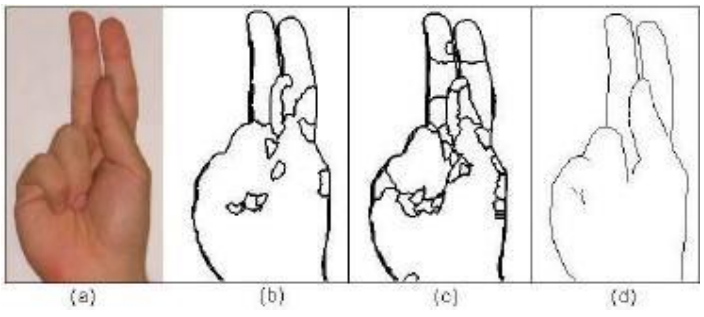

Fig 2.2 :- Image capture mechanism

Skin coloured pixels in the input image are identified by applying a thresholding technique based on the skin colour distribution in YCbCr colour space. The colour of each pixel in the image is set to black or white according to the values of Y, Cb and Cr components. If the Y, Cb and Cr values of a pixel are within a predefined range of skin colour, set that pixel as white otherwise black. Thus, a pixel is classified as belonging to skin if it satisfies the following relation:

$$75 < Cb < 135 \text{ and } 130 < Cr < 180 \text{ and } Y > 80.$$

After image segmentation, we get a binary image containing the handshape representing a particular sign. In order to classify this image, we need to extract certain features of that image. Shape is an important visual feature of an object. Many different methods are available for the representation and description of a particular shape. In this work we propose a new feature for shape representation. The proposed shape feature is derived from the distance transform of the binary image.

Distance transform is a derived representation of an image which is normally applied to binary images. It is also known as distance map or distance field. To apply distance transform on an image, it should be first converted to binary form. A binary image contains object pixels and non-object pixels. Distance transform of a binary image gives another image of the same size where each pixel value is replaced by the minimum distance of that pixel from its nearest background pixel. So, the distance transform of a binary image gives a grayscale image where the gray scale intensity of the foreground region corresponds to the distance from the closest boundary pixel. The three different distance measures used for finding the distance transform of an image are Euclidean, city block, and chessboard. The Euclidean distance transformation is invariant to rotation of the image. So, it is the most commonly used measure for finding the distance transformation, but it involves time consuming calculations such as square, square root and the minimum over a set of floating-point numbers.

There are many techniques to obtain Euclidean distance transform of an image. Most of these methods are either inefficient or complex to implement. Normally the Euclidean distance transform is computed on the basis of a mathematical morphological approach using gray scale erosions with successive small distance structuring elements by decomposition.

The squared Euclidean distance transform is calculated by using a squared Euclidean distance structuring element. The Euclidean distance transform is obtained by applying a square root operation over the squared Euclidean distance transform matrix.

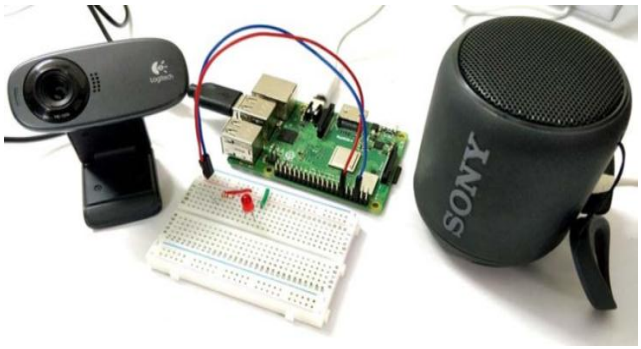

Fig 2.3: Model

- **PERFORMANCE ANALYSIS**

A series of experiments are being conducted to evaluate the system's returns and limitations. In a test done for a sample of 100 spells for different signs. Accuracy up to 90 percent has been achieved. The below given tabular column shows a few random tests

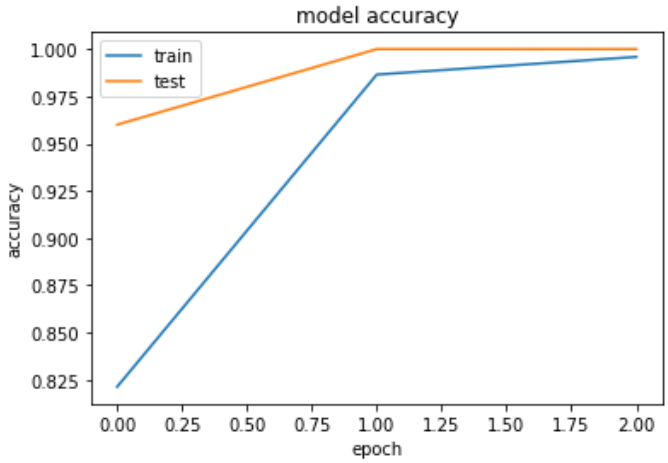

Figure 3.1:- Accuracy

- **Components**

1.) RASBERRY PI :-

Fig 4.1:- Raspberry PI

The Raspberry Pi is a credit card sized single-board computer with an open-source platform that has a thriving community of its own, similar to that of the Arduino. It can be used in various types of projects from beginners learning how to code to hobbyists designing home automation systems.

The Raspberry Pi Model B features:

- More GPIO
- More USB
- Micro SD
- Lower Power Consumption
- Better Audio
- Neater Form Factor

Fig 4.2:- Pin diagram of Raspberry Pi

This higher-spec variant increases the Raspberry pi GPIO pin count from 26 to 40 pins. There are now four USB 2.0 ports compared to two on the Model B. The SD card slot has been replaced with a more modern push-push type micro SD slot. It consumes slightly less power, provides better audio quality and has a cleaner form factor.

2.) CAMERA (WEBCAM) :-

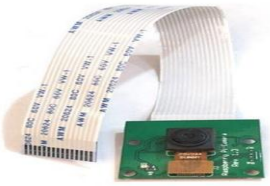

Fig 5.1: Camera

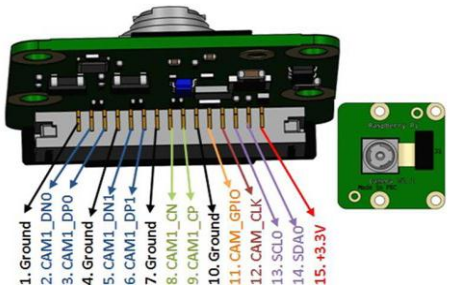

Fig 5.2: Pin diagram of Camera

Sr.no	Pin Name	Description
1	Ground	System Ground
2,3	CAM1_DN0, CAM1_DP0	MIPI Data Positive and MIPI Data Negative for data lane 0
4	Ground	System Ground
5,6	CAM1_DN1, CAM1_DP1	MIPI Data Positive , MIPI Data Negative for data lane 1
7	Ground	System Ground
8,9	CAM1_CN, CAM1_CP	These pins provide the clock pulses for MIPI data lanes
10	Ground	System Ground
1	Ground	System Ground
2,3	CAM1_DN0, CAM1_DP0	MIPI Data Positive , MIPI Data Negative for data lane 0
4	Ground	System Ground
5,6	CAM1_DN1, CAM1_DP1	MIPI Data Positive , MIPI Data Negative for data lane 1
7	Ground	System Ground

v. SOFTWARE REQUIREMENT:-

RASPBERRY PI IDE (PyCharm):

PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the Python language. It is developed by the Czech company JetBrains.It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django as well as Data Science with Anaconda.

PyCharm is cross-platform, with Windows, macOS and Linux versions. The Community Edition is released under the Apache License, and there is also Professional Edition with extra features – released under a proprietary license. PyCharm - Best Integrated Development Environment as it gives possible suggestions at the time of typing code snippets that makes typing feasible and fast.

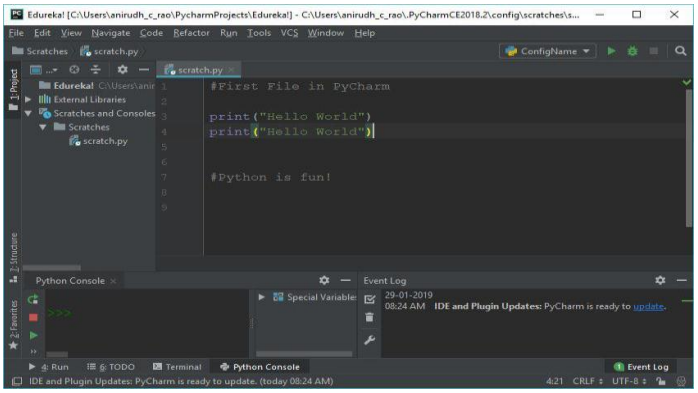

Fig 6: Raspberry Pi IDE (PyCharm)

CONCLUSIONS

The paper is not only aimed at converting the sign language into voice, it's well known that inability to speak and hear is one major challenge for human race.

To overcome these disabilities there are a lot of research and development going in different fields. The paper is aimed to minimize the major complexions in the system for further

extensions, the sensor comes with the feature of face recognition and voice recognition and therefore the next phase of the project would be adding face recognition for capturing the expressions which in turn increases the productivity of the application by adding a little more accuracy.

Also for people with partial voice disabilities the speech recognition system will do the further enhancement in speech systems for the disable people.

REFERENCES

[1] Real-Time Sign Language Gesture (Word) Recognition from Video Sequences Using CNN and RNN, Department of Computer Engineering, Jamia Millia Islamia,New Delhi,India 2018.

[2] Koller, O, Zargaran, O, Ney, H and Bowden, R (2016) Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition In The British Machine Vision Conference (BMVC) 2016, 2016-09-19 - 2016-09-22, York.

[3] T. Starner and A. Pentland, "Real-time American sign language recognition from video using Hidden Markov Models", Technical Report, M.I.T Media Laboratory Perceptual Computing Section, Technical Report No. 375, 1995.

[4] Deng-Yuan Huang1, Wu-Chih Hu2, Sung-Hsiang Chang, "Vision- based Hand Gesture Recognition Using PCA + Gabor Filters and SVM" ,2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing.

[5] Rini Akmeliawati, Melanie Po-Leen Ooi and Ye Chow Kuang, "Real- Time Malaysian Sign Language Translation Using Colour Segmentation and Neural Network", IEEE on Instrumentation and Measurement Technology Conference Proceeding, Warsaw, Poland 2006, pp. 1-6.

[6] Nadia R. Albelwi, Yasser M. Alginahi, "Real-Time Arabic Sign Language (ArSL) Recognition "International Conference on Communications and Information Technology 2012.

[7] Stephan Liwicki, Mark Everingham, "Automatic Recognition of Finger spelled Words in British Sign Language", School of Computing University of Leeds.

[8] Azadeh Kiani Sarkalehl, Fereshteh Poorahangaryan, Bahman Zan, Ali Karami, "A Neural Network Based System for Persian Sign Language Recognition" IEEE International Conference on Signal and Image Processing Applications 2009.

[9] Al-Jarrah, A. Halawani, "Recognition of gestures in Arabic sign language using neuro-fuzzy systems," The Journal of Artificial Intelligence 133 (2001) 117–138.

[10] Incertis, J. Bermejo, and E. Casanova, "Hand Gesture Recognition for Deaf People Interfacing," The 18th International Conference on Pattern Recognition, 2006 IEEE.

[11] R. Rokade, D. Doye, and M. Kokare, "Hand Gesture Recognition by Thinning Method" International Conference on Digital Image Processing (2009).

[12] R. Feris, M. Turk, R. Raskar, K. Tan, and G. Ohashi. "Exploiting depth discontinuities for vision-based fingerspelling recognition". In IEEE Workshop on Real-time Vision for Human-Computer Interaction, 2004.