ISSN 2395-1621

Sign Language Interpreter with voice based output using CNN algorithm

Rucha Kolte, Sonali Nikam, Shradha Malwad, Komal Borhade,
Prof. Madhav Ingle

kolte.rucha1996@gmail.com, sonaliwawarhire@gmail.com, shradhamalwad1198@gmail.com, borhadekomal61@gmail.com, ingle.madhav@gmail.com

Department of Computer Engineering
JSCOE, JSPM's, Hadapsar, Pune.

ABSTRACT

Addressing the issues of People with Hearing and Vocal Impairment through a single aiding system is a tough job. A lot of work in modern day research focuses on addressing the issues of one of the above challenges but not all. The work focuses on finding a unique technique based on the machine learning that aids the mute by letting them hear what is represented as text and its sound. The proposed system achieved the technique that takes the sign image through a web camera and applies to the image processing then analysis what exactly want to the mute people at end the text available to voice signals. All these three solutions were modulated to be in a single unique system. All these activities are coordinated using the Ubuntu system using python. The vocally impaired people are helped by the process in which the image to text and text to speech is given using machine learning. It would take sign language as input which would display result not only in the form of text, it would display the corresponding image.

Keywords: Hand sign detection, Sign detection, machine learning, Image Processing.

ARTICLE INFO

Article History

Received: 20th June 2020

Received in revised form :

20th June 2020

Accepted: 23rd June 2020

Published online :

23rd June 2020

I. INTRODUCTION

Approximately 285 million people are judged to be visually impaired worldwide in which 39 million are blind and 246 are said have low vision. Approximately 90% of this world's vocally impaired is from the dispirited income people and 82% of people living with blindness aging persons and above. The numbers of people visually impaired from eye related diseases have been brought down in the past 20 years according to global estimated work. In which 80% of all visual restitution can be prevented or cured. India is considered to be the home for the world's largest act of blind people. In this world, about 37 million are blind, in which 15 million are from India. There are so many researches have been getting along in this universe, but the visual impairment could not be broken for good. In lodge to facilitate these people we have developed the assistive device for blind people who does not want the assistance of other neighbors. Dumber people can simply tilt the message by sign language which could not be understandable by other people. In resolving these difficulties with visually and vocally impaired people we have used the sign image to the system and generate the voice. By this device we provide the solution for blind, deaf and dumb people. For blind people the image is converted to voice by using Tesseract software, the deaf people received their content by message as soon as the opposite person speaks out it displayed as a message. The dumb persons conveyed their message through text instead of sign language which is delivered via e speak. We have provided necessary steps to resolve the problems of those masses. The motivation for a hand gesture recognition is to assist handicapped users. We can provide quality assistance to the physically challenged users, also for senior citizens by devising Image Processing techniques. It is manual operation. Persons actions are difficult to understand. Sometime person action or gestures are difficult to recognized as it is tough job. Then in that case communication getting difficult and more inconvenient. Conveying information to be take more time. It is difficult as well as very time consuming. These are main problems which create disturbance in communication.

II. LITERATURE SURVEY

The literature survey is considered as a part of the work. It interference the queries related the improvement of work

already done and clearly outline the development of the research projects.

In this paper various technological developments that have improved the interaction among the deaf and nondeaf have been discussed. Two different strategies are drawn out in this paper. The first is using wearable communication devices such as gloves or using a touch screen device to communicate[1].

This paper focuses on bare hand gesture recognition system by proposing a scheme using a database-driven hand gesture recognition based upon skin color model approach and thresholding approach along with an effective template matching with can be effectively used for human robotics applications and similar other applications.. Initially, hand region is segmented by applying skin color model in YCbCr color space[2].

In this paper, author talk about the use of selective panel that depicts what a concern person want to communicate through various set of images stored in it and then speak out using the device[3].

The proposed method uses a hierarchical CRF to detect candidate segments of signs using hand motions, and then a BoostMap embedding method to verify the hand shapes of the segmented signs. The proposed method could recognize signs from signed sentence data at a rate of 90.4%[4].

This paper presents the prototype of a low cost smart glove to improve the mobility of the visually impaired people. The glove is equipped with rangefinders to explore the surroundings: it provides a vibro-tactile feedback on the position of the closest obstacles in range by means of vibration motors[5].

This paper presents an approach to extract features by using Mel Frequency Cepstral Coefficients (MFCC) from the speech signals of isolated spoken words. And, Hidden Markov Model (HMM) method is applied to train and test the audio files to get the recognized spoken word. The speech database is crea ted by using MATLAB[6].

In this study the letters and the numbers of the Turkish Sign Language are detected by using a data glove which is composed of strain sensors, a gyroscope and an accelerometer. The sensors used on the glove are characterized, the design of the glove and the classification of the sensors' data are explained[7].

In this paper author method uses a hierarchical CRF to detect candidate segments of signs using hand motions, and then a BoostMap embedding method to verify the hand shapes of the segmented signs Experiments demonstrated that the proposed method could recognize signs from signed sentence data at a rate of 90.4%[8].

III. PROBLEM STATEMENT

In existing system the module was developed for dumb person using flex sensor, there user hand is attached with the flex sensors. On this module the flex sensor reacts on bend of each finger individually. By taking that value controller starts to react with speech, each flex sensor holds unique voice stored in APR Kit and for each sign it will play unique voice. And in other existing system, the work is done only for some alphabets and not for the words or sentences, and accuracy obtained is very low. The work focuses on finding a unique technique that aids the mute by letting them hear what is represented as text and it is achieved by the technique that captures the hand sign image through a web camera and converts the text available to voice signals.

IV. DESIGN AND METHODOLOGY

A. Hardware & Software used:

- Camera (USB 2.0 webcam)
- Raspberry pi
- Python 2.7 or above
- Opencv
- Tensorflow

B. Design structure:

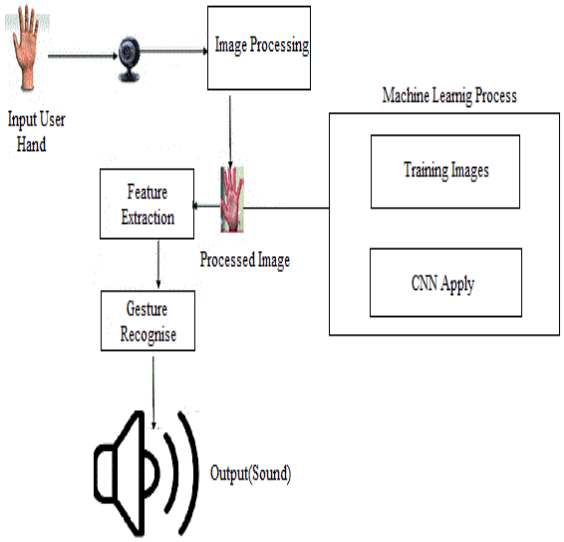

Fig 1. System architecture

C. Algorithm Process:

CNN ALGORITHM:

Step 1: Convolution Operation

Here are the three elements that enter into the convolution operation:

1. Input image
2. Feature detector
3. Feature map

Step 1(b): ReLU Layer

The reason we want to do that is that images are naturally non-linear.

When you look at any image, you'll find it contains a lot of non-linear features (e.g. the transition between pixels, the borders, the colors, etc.). The rectifier serves to break up the linearity even further in order to make up for the linearity

that we might impose an image when we put it through the convolution operation.

Step 2: Pooling

Again, max pooling is concerned with teaching your convolutional neural network to recognize that despite all of these differences that we mentioned, they are all images are same. In order to do that, the network needs to acquire a property that is known as "spatial variance." This property makes the network capable of detecting the object in the image without being confused by the differences in the image's textures, the distances from where they are shot, their angles, or otherwise.

Step 3: Flattening

This will be a brief breakdown of the flattening process and how data move from pooled to flattened layers when working with Convolutional Neural Networks.

Step 4: Pooling

What happens after the flattening step is that you end up with a long vector of input data that you then pass through the artificial neural network to have it processed further which is called pooling.
Types of pooling: Mean, Max, Sum

Step 5: Full Connection

In this part, everything that we trained throughout the section will be merged together. By learning this, you'll get to envision a fuller picture of how Convolutional Neural Networks operate and how the "neurons" that are finally produced learn the classification of images.

Step 6:Summary

In the end, it will wrap everything up and give a quick recap of the concept covered in the training.

Step 7: SoftMax& Cross-Entropy

Optimization Functions for CNN model. To calculate final accuracy and losses.

CNN IN Our Project:

1. Classify dataset under labeled folders such as had signimages
2. Read dataset
3. Read features of all images and label (here name of dataset folder) of it
4. Store it in model file
5. Get input image
6. Read features of input image
7. Compare features of stored features
8. Show label as prediction of nearly matched features.

V. RESULT ANALYSIS

Fig 2. Camera Captured Greyscale Image

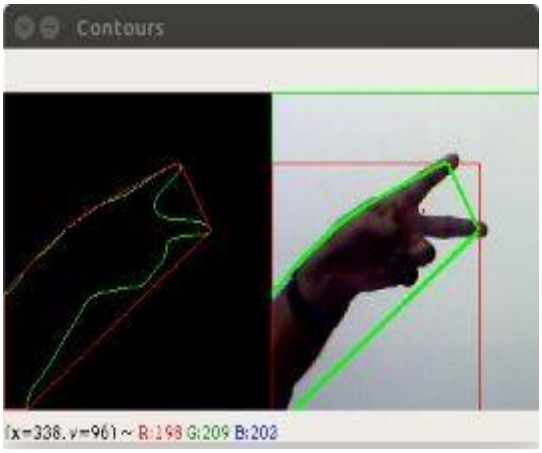

Fig 2. Contours Image

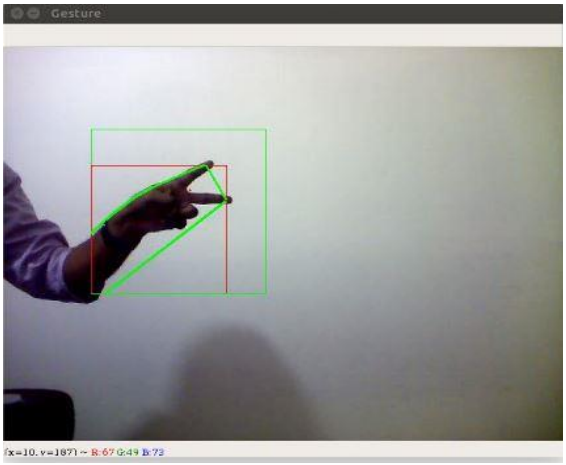

Fig 4. Gesture analysis image

VI. CONCLUSION

The implementation of the proposed system aims to translate gestures into speech (voice). The scope of the project is to enhance the recognition capability for various lightning conditions and achieving more accuracy. Implementing and identifying the more number of gestures. The proposed system achieved the technique that takes the sign image through a web camera and applies to the image processing then analysis what exactly want to the mute people at end the text available to voice signals. All these three solutions were modulated to be in a single unique system. All these activities are coordinated using the Ubuntu system using python. The vocally impaired people are

helped by the process in which the image to text and text to speech is given using machine learning. It would take sign language as input which would display result not only in the form of text, it would display the corresponding image. So By performing this project we were able to detect the hand gesture regonistion using Machine Learning and hence this project will help the mute and deaf people to communicate easily and express their thoughts and feeling to their loved once.

VII.REFERENCE

[1] Sunitha K. A, AnithaSaraswathi.P, Aarthi.M, Jayapriya. K, Lingam Sunny "Deaf Mute Communication Interpreter-A Review", 2016.

[2] Mandeep Kaur Ahuja, Dr.Amardeep Singh, "Hand Gesture Recognition Using PCA", 2015.

[3] Kunal Purohit, Prof.Kailash Patidar, Mr.Rishi Singh Kushwah, "A Wearable Hand Gloves Gesture Detection based on Flex Sensors for disabled People, 2017.

[4] Hee-Deok Yang Sign Language Recognition with the Kinect SensorBasedonConditional Random Fields", 2015.

[5] Giuseppe Bernieri, Luca Faramondi and Federica Pascucci, "A Low Cost Smart Glove for Visually Impaired People Mobility", 2015.

[6] Su Myat Mon, HlaMyoTun, "SpeechTo-Text Conversion (STT) System Using Hidden Markov Model (HMM)", 2015.

[7] Fatih Kaya, Ahmet Furkan Tuncer, ŞölenKumbayYıldız, "Detection of the Turkish Sign Language Alphabet with Strain Sensor Based Data Glove", 2018.

[8] Hee-Deok Yang, "Sign Language Recognition with the Kinect Sensor Based on Conditional Random Fields" 2014.

