ISSN No:-2456-2165

# Sign Conversion for Hearing Impaired People

Slomo A Thomas, Rahul Ajith, Sebin Skariah, Visakh S Nair
Sarju. S, Assistant Professor
Department of Computer Science and Engineering
St Joseph's College of Engineering and Technology Palai, Kerala, India

**Abstract:- Sign languages are languages used by deaf community in which different means of hand sign expression are used for communication. There are about 10 to 15 million deaf in india. In India, there is no universal sign language for the deaf community to express themselves . A person who is deaf and uses only this gestural sign language for communication will find it hard to converse with the person who does not know sign language. This causes various problems. Though there exist many sign languages, the common people are not aware of sign languages. Hence communicating with deaf and dumb people becomes more difficult. There is a need for a middle person to translate what they want to express. Our work aims to improve the communication with the deaf. The proposed concept(sign language conversion) is a breakthrough for helping the deaf community. The system uses Convolutional Neural Network(CNN) algorithm. The algorithm first acquires the input from the user and preprocessing of hand signs or gestures is performed to capture and match the hand signs. Using a camera, the video is captured and the gestures are extracted using feature extraction. The captured gestures are processed and then converted into text and voice**.

***Keywords:-*** *Convolutional Neural Network(CNN), sign language conversion.*

## I. INTRODUCTION

One of the greatest difficulties faced by the deaf community is the hearing culture that treats them. Sign languages are the languages that use the hand signs and gestures to express meaning [1]. Whenever the deaf community exists, sign languages have been the handy means of communication between them. It is also used by the people who are unable to speak

This paper utilizes a strategy for perceiving gesture based communication employed in a web-based application The system is mainly used in the communication between deaf and the normal people. Ordinary individuals are totally ignorant of the communication via gestures that the deaf individuals use [3]. Henceforth correspondence with them is beyond the realm of imagination except if they took appropriate training to get familiar with this gesture based communication.

The preferred system is to help the normal people for a good understanding with the deaf community by converting the sign languages into text or voice [2]. The image recognition of the hand gestures is done by using a CNN(convolutional neural network). Convolutional neural network is one of the deep learning algorithms.

The proposed method of sign language recognition will definitely be useful for the effective communication between deaf and the normal people. The voice output incorporated with the application will give the deaf users an exquisite feeling, they will no longer be deaf. This paper strives to provide an efficient camera based sign language conversion system.

## II. RELATED WORKS

### *A. Smart glove interpretation*

The existing system for sign language recognition involves the use of glove sensors. The input signs given by the user are acquired and identified using the glove sensors [4]. These sensor gloves have multiple sensors inserted in them. They are mainly used for tracking the location and movement of the fingers and the palm. One major advantage of using sensor gloves is that it can detect and identify hand signs without being affected by external disturbances due to change in light, electric field or magnetic field [5]. The major problem lies in its complexity of using sensors for gesture recognition. The user interaction with the system is done by an interface called Kinetic user interface (kui) by detecting the motion of the object under consideration, in our case the human hands. The user is relied upon to be wearing the sensor gloves whenever he needs to give an input.

### *B. Practikality*

The website application is to empower people with disabilities. The machine learning-based practikality application provides a fully-interactive mechanism for the deaf, blind, and mute to interact with ease. This arrangement can be utilized by the mute to convert over content into voice [6]. The client needs to enter in the message that the person in question needs to convey. The application changes over the information which can be either words or into sound, along these lines encouraging correspondence between the client and the beneficiary. This solution does not provide a real time image, it just takes videos from youtube and conveys it.

## III. PROPOSED METHOD

A sign language is a language that can be used for communication using different hand signs . It is based on indian hand signs. Also voice out for each letter identified from the hand sign. Convolutional Neural Network(CNN) is used for the recognition of hand signs. 24 Letters are trained and recognized in our system.

➢ *Convolutional Neural Network(CNN)*

Convolutional neural networks (CNN) is an emerging artificial neural network structure as it is used in both image recognition as well as in speech recognition. CNN is applicable in video analysing, drug analysis, natural language processing [7]. Based on the deep learning of the CNN method the feature extraction and classification process can automatically learn the appropriate characteristics and classification in the image recognition field. The preprocessing required in a CNN is much lower as compared to other classification algorithms.

The entire project is divided into three sections: image pre-processing, feature extraction, classification.

- *Image processing*

In the image processing phase the captured video is converted to frames for feature extraction. The different image preprocessing stages performed tends to improve the overall performance of the system by modifying the image or video inputs [8]. Medium filter and gaussian channels are a portion of the generally utilized systems to diminish commotions in pictures or video gained. The removal of unwanted information is done by morphological operation, which is widely used. In the first threshold the input image is converted into a binary image followed by median and gaussian filters which are used to remove noises. Next is the morphological operation which removes unwanted information as mentioned above [8]. The captured images are scaled to a lower resolution prior to the succeeding stages.

This entire process is depicted in Fig 1. The overall computational efficiency of the system is improved by reducing the resolution of the image. Different images captured under different environments have different illumination and brightness levels. In order to avoid this histogram equalization is used to improve or enhance contrast to a uniform illumination and brightness levels.

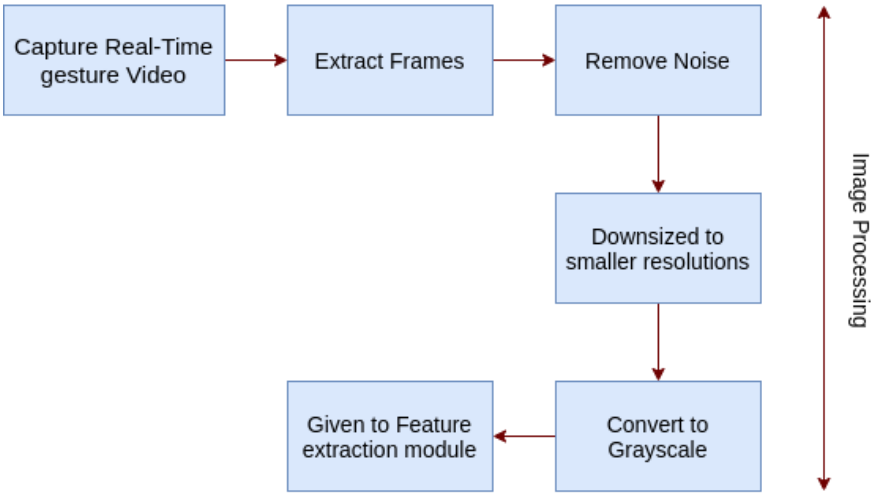


Fig 1:- Image Processing

- *Feature extraction*

Feature extraction is the change of fascinating pieces of information into sets of minimal component vectors. It involves in reducing the number of resources to describe a large set of information [9]. In this phase the processed image is given as input and from it, data like hand signs, shapes, edges etc is detected and extracted. The highlights or data separated should contain applicable data from the hand input to a minimal variant [9]. The principal objective of feature extraction is to extract or retrieve data that matters. In this case the hand signs or gestures are extracted from the input and the data is used in a lower dimensionality space. The Feature extracted or data is given to the classification module for identifying the signs and converting them to text/label and finally as a voice output. Fig 2 shown below depicts the basic feature extraction process in sign interpretation.

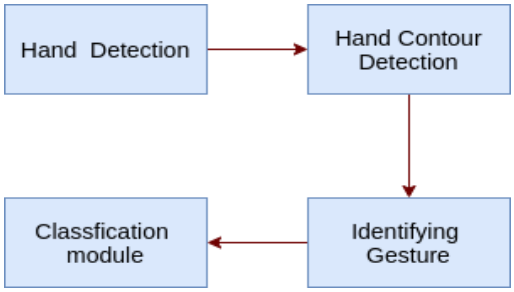


Fig 2:- Feature Extraction

- *Classification*

Classification can be classified into supervised and unsupervised machine learning methods. Supervised is a method that shows the framework to perceive certain examples of info information, which are then used to foresee future information [10]. Directed learning takes in a

ISSN No:-2456-2165

lot of known preparing information and it is utilized to gather a capacity from named preparing information. An unsupervised learning is utilized to draw deductions from datasets with input information with no marked reaction. Since no named reaction is taken care of into the classifier, there is no prize or punishment weightage to which class the information should belong.Image arrangement alludes to a procedure in pc vision that can order a picture as per its visual substance.

## IV. SOFTWARE IMPLEMENTATION

The goal of CNN is to learn the features present in the data with higher order using contortion [11]. The identification of objects such as images is done with the help of CNN architecture. This includes faces, street signs and other aspects of the visual data. The core building block of CNN is the convolutional layer. Convolutional layers revise the input using a neuron patch that is connected locally from the layer prior to it. The inner product will be calculated by using the layer between the region of the neurons present in the input layer and the weights to which they are locally connected present in the output layer. Convolution is said to be a mathematical operation that describes the rule for combining two sets of data. The convolution operation takes the input and a convolution filter is applied to it. The purpose of other feature extraction layers is to minimize the size of the result obtained by convolutional layers [11]. After convolution, the max-method will be used over a region with some specific size for subsets of the feature map. The reduction of characterization of data is done using the pooling layers. It progressively reduces characterization of data over the network and control over-fitting. The pooling layer works in an independent manner on every single depth slice of the input data. The pooling layer uses the max() operation that helps to change the size of the given input data spatially [5]. This is said to be max-pooling. The filters on the input data are used for performing down-sampling in this layer.

The different features of classification determine the scores of fully connected layers or output [12]. The output volume has a dimension of [1 9 1 9 n], where n represents the number of output classes to be evaluated. All the neurons in the preceding layer have different sets of weight which are connected with the output neuron in each layer. Furthermore, the fully connected layer is a set of convolutions wherein each feature map [12] is connected with each field of the back to back layer and filters comprise of a similar size as that of the input image. By changing the number of pooling and convolution layers the system is checked. In order to improve the adequacy of the outcomes, one more layer, i.E., Dropout layer, is likewise included in the proposed approach. Adding a Dropout layer is a type of regularization procedure which is used to overlook haphazardly chosen neurons at the time of training. It also helps in diminishing the odds of over-fitting.

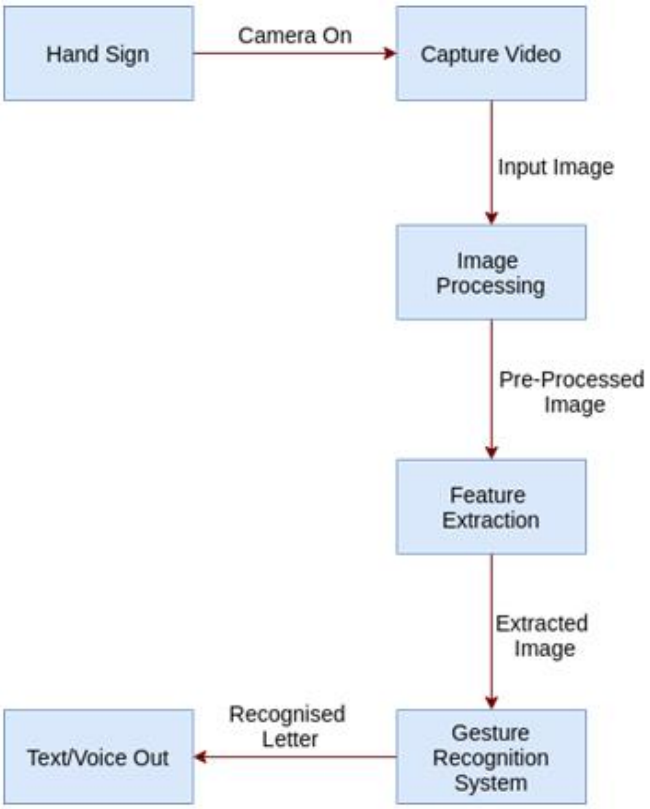


Fig 3:- System Design

➢ *Sign to text or voice conversion*

First is the system initialization( all the libraries are called), where all the libraries are called and the system gets activated. Once the system is active, then the camera is turned on to capture the video. The captured video is converted to frames. These frames are then given for image processing. During image processing the errors are removed. For better optimization purposes the image is converted to grayscale as pixel size. Next is the feature extraction part where all the features from images is identified and extracted. The extraction is supplied to Gesture Recognition System. Then it is fed to the model and the model is trained.

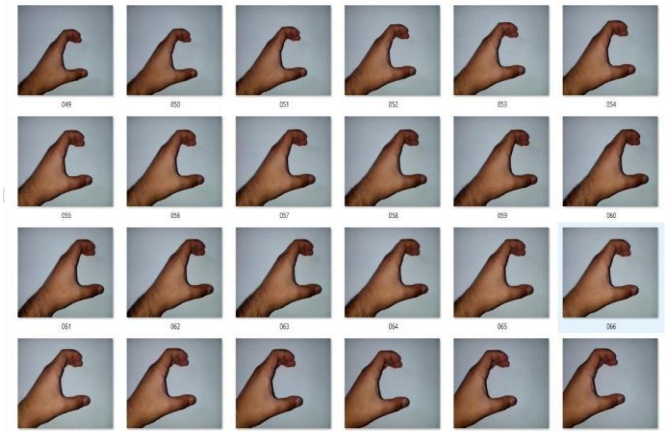

Fig 4:- Datasets

ISSN No:-2456-2165

After training the input is provided by the user via the training model. The identified signs from the input are converted into label's or text output as shown in Fig 3. Then labels are then converted to voice output as shown in Fig 3.

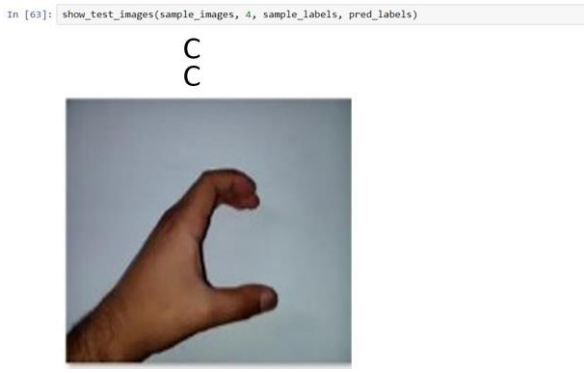


Fig 5:- Identified sign is converted to text

The entire process of sign to text or voice conversion is shown by means of a flowchart in Fig 6.

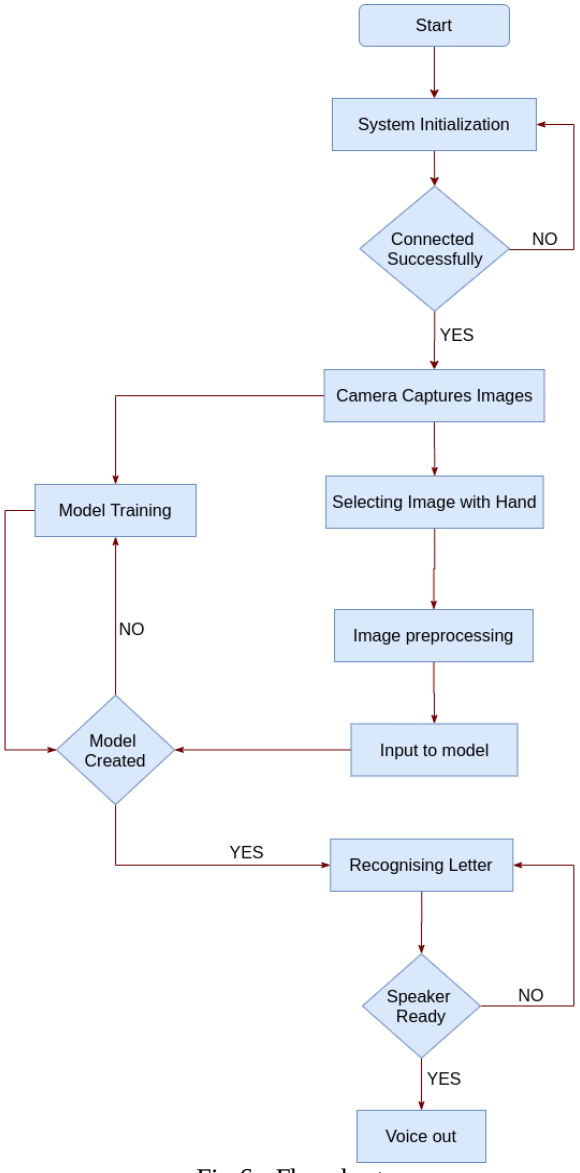


Fig 6:- Flowchart

## V. RESULTS & CONCLUSION

Fig 5 shows the result that has been obtained(text output). The real time sign conversion system has produced a better accuracy that is somewhat higher than existing systems and has a more user friendly approach as it is capable of converting the sign languages to text as well as voice format. The identified sign form input is converted into text output and these label's are then converted to voice.

The framework for perceiving real-time Sign Language depicts a noteworthy job in upgrading easy going correspondence among individuals with hearing disabilities and ordinary people. Likewise, for high dimensionality datasets, the vast majority of the customary calculations endure. Subsequently it is intended to broaden the framework by joining Convolutional Neural Networks (CNN) to catch the spatial and transient highlights.

## ACKNOWLEDGMENTS

This project is developed by final year undergraduate students in APJ Abdul Kalam Technological University, Kerala under the supervision of Asst.Prof. Sarju. S

## REFERENCES

[1]. Umang Patel, Aarti G. Ambekar, " Moment Based Sign Language Recognition For Indian Languages" .
[2]. Joyeeta Singha, Karen Das, " Recognition of Indian Sign Language in Live Video ".
[3]. Dipti Jadhav, Amiya Tripathy, " Gesture Aided Speech for Deaf and Mute ".
[4]. Abhinandan Das, Lavish Yadav, Mayank Singhal, " Smart glove for Sign Language communications" .
[5]. Nitipon Navaitthiporn, Preeyarat Rithcharung, Phitnaree Hattapath, C. Pintavirooj " Intelligent glove for sign language communication ".
[6]. https://coolblindtech.com/students-develop-practikality-app-for-deaf-blind-and-non-verbal-people-to-communicate/
[7]. Lakshman Karthik Ramkumar, Sudharsana Premchand, Gokul Karthi Vijayakumar " Sign Language Recognition using Depth Data and CNN ".
[8]. Omkar Vedak, Prasad Zavre, Abhijeet Todkar, Manoj Patil " Sign Language Interpreter using Image Processing and Machine Learning ".
[9]. Gaurav Kumar, Pradeep Kumar Bhatia, "A Detailed Review of Feature Extraction in Image Processing Systems" .
[10]. Zuzanna Parcheta, Carlos-D. Mart´ınez-Hinarejos " Sign Language Gesture Classification Using Neural Networks ".
[11]. Gongfa Li, Heng Tang, Ying Suu, Jianyi Kong, " Hand gesture recognition based on convolution neural network" .
[12]. Nadia Jmour, Sehla Zayen, Afef Abdelkrim " Convolutional neural networks for image classification " .