ISSN No:-2456-2165

Sign Language Recognition System with Speech Output

Vinayak S. Kunder
Dept. Student of Computer Engineering
Pillai HOC College of Engineering and Technology
Rasayani, India

Aakash A. Bhardwaj
Dept. Student of Computer Engineering
Pillai HOC College of Engineering and Technology
Rasayani, India

Vipul D. Tank
Dept. Student of Computer Engineering
Pillai HOC College of Engineering and Technology
Rasayani, India

Abstract - Indian deaf and mute community is troubled by a lack of capable sign language volunteers and this lack is reflected in the relative isolation of the community. There have been many techniques used to apply machine learning and computer vision principles to recognize sign language inputs in real-time. CNN is identified as one of the more accurate of these techniques. This project aims to use CNN to classify and recognize sign language inputs produce text-to-speech outputs. In doing so, sign language speakers will be able to communicate to larger audiences without worrying about the speed a human interpreter. OpenCV is used for live detection of the hand gestures performed by the user. We are using Indian sign language to train the model. After recognition of a particular gesture, it will convert the predicted text to speech so that other person can hear what the user want to convey the message.

Keywords:- *Sign-language, OpenCV, Gestures, Convolutional Neural Networks(CNN).*

I. INTRODUCTION

Communication is one of the reasons for the rapid progress that mankind has been making in the fields of art, architecture, music, sports, science, technology, drama and so on. It lies at the root of progress and is the most essential tool for survival as a group and society. Most of us are fortunate enough to have the ability to communicate verbally, while others are not so fortunate being vocally challenged and hearing impaired. Sign language comprises lots of complex hand movements, and every tiny posture can have a variety of possible meanings[6].The way they communicate has been evolving from pointing out in directions to certain objects to having a specific set of gestures to convey meaning in the form of sign language. The people who are vocally challenged and hearing impaired use different forms of sign language all over the world. This helps them get the education and learning equivalent to their more gifted peers. While intellectually unaffected by their ability to speak or hear, the sign language community all over the world is still functioning within relative isolation. The Indian Sign Language has about 2.7 million speakers and roughly about 300 known volunteering interpreters. It is not feasible for such a low number of interpreters to work with that many speakers. It becomes very challenging for a sign language speaker to interact with someone outside the sign language community since there is a lack of volunteering interpreters or any other facilities which can support a conversation between these two individuals. Greater challenges lie when a sign language speaker has to interact with a large audience. Sensor-based methods provide feasible solutions, but wearing extra equipment on hands is inconvenient to people[7].

Roughly 28 million people in India suffer from some level of hearing loss, says research by Centres for Disease Control and Prevention. A small percentage of these people may not use sign language at all, instead using methods like lip reading. A person who communicates using lip reading can have a more active participation in real world conversations. On the other hand, those who only use sign language are limited to carrying out conversations with fellow sign language speakers only. Therefore, they require interpreters to engage fully in real world conversations.

II. LITERATURE SURVEY

A. Suharjitoa, Ricky Anderson , Fanny Wiryana , Meita Chandra Ariesta , Gede Putra Kusumaa, Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output, 2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI 2017, Bali, Indonesia.: This paper studies the various phases of our intended application such as data acquisition, image processing, other suggested classification methods and their comparison with Convolutional Neural Networks. The accuracy result for classification of CNN has shown to be 94.2% which encouraged us to do the same with similar results in our implementation of CNN.

B. Pratibha Pandey, Vinay Jain, "Hand Gesture Recognition for Sign Language: A review,", IJSETR, Vol.4, Issue 3, March 2015: The objective of this paper is to highlight widely effective methods of capturing gestures which have been fundamental in the recent past.

C. Manisha U. Kakde, Mahender G. Nakrani, Amit M Rawate, “A Review Paper on Sign Language Recognition For Deaf And Dumb People using Image Processing,” ,IJERT, ISSN:2278-0181, Vol.5 Issue 03, March 2016: This paper lists current most popular methods of sign acquisition.

D. Shreyasi Narayan Sawant, M. S. Kumbhar, “Real Time SignLanguage Recognition using PCA”, 2014 IEEE conference on Advanced Communication Control and Computing Technologies: The results of this paper

demonstrate that by using simple hand gestures the corresponding letters can be predicted and obtained as output.

III. PROPOSED WORK

The input will be a webcam live feed of a human speaking in sign language. The waist-up portion will be within the frame to maximize the variety of signs taken as input. The challenge we face is the application of a suitable method in computer vision that can detect the smallest discrepancy between similar signs to avoid inaccurate translation in real time. We also need to train a separate dataset for random gestures and movement to ensure they are not translated to convey meaning. Next, we need to instantly display captions which are a result of the translation of sign language inputs, on the screen to give extra assistance to the audience. Since our final goal is to give a sound output for the aforementioned translation of sign language, we need to employ a suitable text-to speech software or a web API that can convert this input into sound output in the most human way possible to avoid a robot-like sound output.

A. Sign Language

Humans are social animals who have communication at the core of their characteristics. Languages have helped us evolve as a species. One such language is the sign language. Sign language is essentially a means of communication used to convey meaning through the use of gestures and symbols. Over the world, there are different versions of sign language depending on that region. [5]. ISL (Indian Sign Language) consists of both isolated sign as well as continuous signs An isolated sign refers to a single hand gesture, and a continuous sign is a moving gesture represented as a series of images. It helps them gain knowledge and also gives them a real chance at having a formal education. In India, the lack of knowledge about sign language has caused speech and hearing impaired people to drop out of the education system rapidly. This is something we aim at changing. The fight to make Indian sign language official has been encouraging lately as 'The First Indian Sign Language Dictionary of 3000 Words' was launched in march 2018. We realised that some level of automation is needed in this role to help the deaf and mute gain access to education and further improve their lives.

B. Computer Vision

Computer Vision is at the crux of our work, it is a technology that allows computers to perceive the outside world through the lens of a camera and make logical interpretations. Computer Vision has evolved rapidly since its applications and potential were noticed shortly after its introduction. Computer Vision aided machine learning is being used to solve a massive number of real world problems such as automated driving, surveillance, object detection, facial recognition, gesture detection and so on. Computer vision and health care has the potential to deliver real value, while the computers wont replace the healthcare personnel, but there is a possibility to improve routine diagnostics that require a lot of time and expertise.In this way the computer vision serve as a helping tool for healthcare. Our work is based on acquiring visual input and processing it further. There are a number of libraries available that facilitate use of computer vision such as OpenCV, PCL (Point Cloud Library), ROS for robotic vision, and MATLAB. We have used OpenCV for its seamless integration into Python programs. Training and classification in OpenCV is done with the help of neural networks, which have shown to have an accuracy of as much as 94.2% in [1]. Neural networks used for computer vision applications require a lot of high-quality data. The algorithms need a lot of data which should be related to the project, this will produce good results. Images are available online in large quantities, but the solution to many real-world problems needs quality labelled training data, this makes it expensive because the labelling has to be done manually.

C. Convolutional Neural Networks(CNN)

A Convolutional Neural Network is one of the types of artificial neural network(ANN)[2] which makes the use of convolutional layers for filtering inputs and extraction of information. The convolution function combines input data also known as feature map with a convolution kernel(filter) to get a transformed feature map. The filters in convolutional layers are transformed according to learned parameters to extract the most relevant information for a given task. Convolutional neural networks make automatic adjustments to find the most suitable features required by as task. The Neural network would filter information related to the shape of an object when assigned an object recognition task but will extract the colour of the animal when assigned an animal recognition task. This is based on the ability of the Convolutional Neural Network to understand that different classes of objects will have different shapes but that different types of animals will likely differ in colour than in shape. Convolutional Neural Networks have various applications which include natural language processing, sentiment analysis, image classification, image recognition, video analysis, text analysis, speech recognition, text classification processing applications, with Artificial Intelligence systems such as virtual assistants, robots, and autonomous cars, drones, and manufacturing machines. CNN has been important in our implementation being useful for training our dataset of multiple gestures. The CNN consists of three layers namely an input layer, output layer and a hidden layer which consists of the convolutional layers, pooling layers, fully connected layers and normalization layers.

D. Technical Details

The need to derive a solution to the problem that is lack of sign language volunteers encouraged us to make a system that can take signs as an input, analyse it with the dataset, generate a string output for a specific sign using it to generate a speech output. We have included functionality to enable users to train their own gestures.

We used Python for the programming part along with OpenCV, and pyttsx library for real-time speech generation

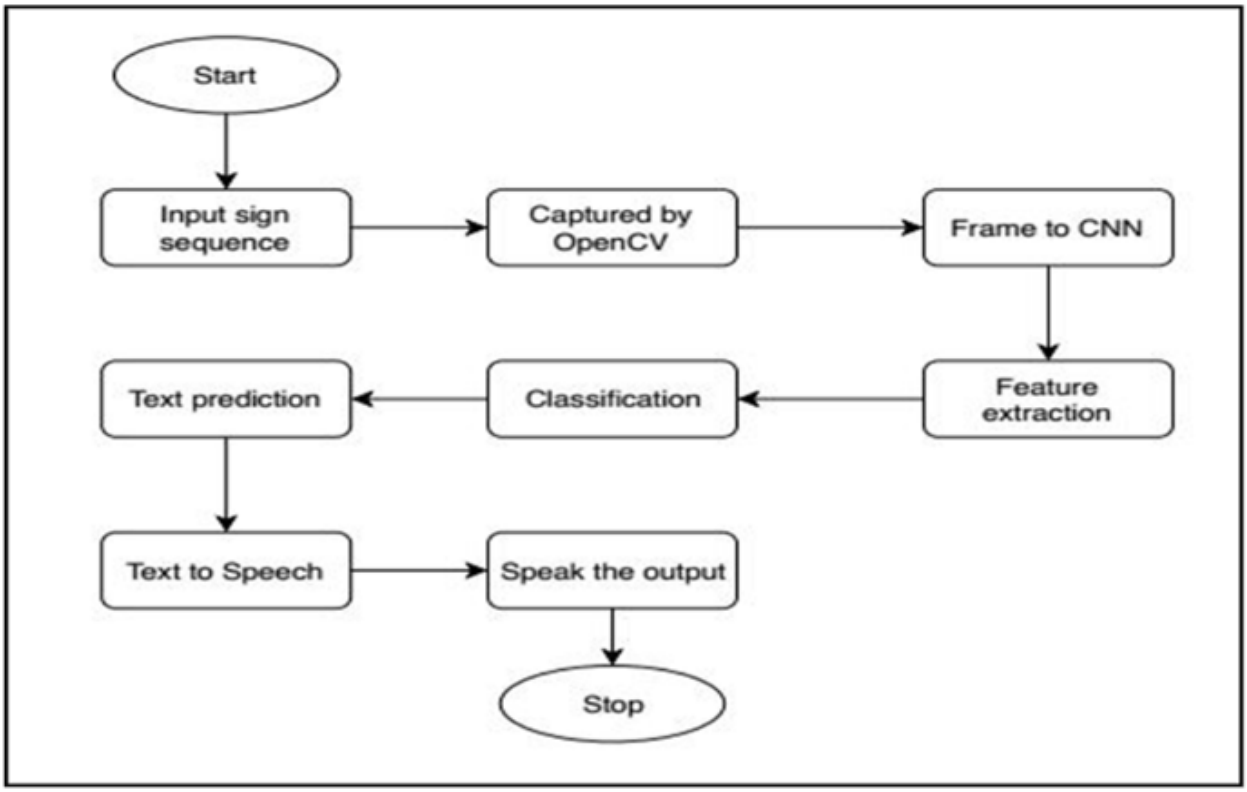

Fig 1

➢ *Data Acquisition*

The form of data that needs to be processed and hence acquired is image data which can easily be captured by a webcam like [2] with the help of OpenCV functions which allow live video to be broken down into single frames representing individual images. [3] have used a 3d Model approach We have captured our own dataset for a particular gesture, so that user can set his own gesture for different words and they do not have to remember the standard gestures. To enhance the specificity of acquired data, we can use skeletonization and background reduction techniques. Skeletonization reduces the main subject of the image into a skeleton-like frame joined at points. Background reduction is used to remove redundant data that is interfering with the subject and causing any inefficiency. These images are converted to numerical form using the numpy library that facilitates mathematical functions finally converting each frame into a matrix containing values for each pixel in the image.

➢ *Training CNN model*

We trained CNN model using Keras deep learning library, Keras is an open source neural network library which uses tensorflow in its background for processing. Keras allows us to easily experiment and test the neural networks models just by tuning some input values. We have built total three convolution layers. [4] Each gesture in the dataset is an image of 240x320 pixels, these images are converted in the form of matrix.The columns of the matrix represents the pixels of the image and range from 0 to 255.

- **First Layer** - The first layer has 32 filters for building the feature maps, each feature map is of 3x3 size. The input shape is set to 64x64 size and the activation function used is RELU (Rectifier Linear Unit). The maximum pooling size is set to 2x2.
- **Second Layer** - This layer is similar to the first layer same filters and feature maps and relu activation function.
- **Third Layer** - In this layer there are 64 filters and 3x3 sized feature map with relu activation function and maxpool size 2x2. After this the flatten layer is added and full connection layer contains two dense layers with 256 filters. The second dense layer contains the softmax function because there are more than two types of outputs to predict.

➢ *Recognition*

There are many different ways for performing image recognition, but many of the top techniques involve the use of convolutional neural networks. The CNN can specifically set for image processing and recognition. Using a combination of techniques such as max pooling, stride configuration and padding, convolutional neural filters work on images to help machine learning programs identify the subject of the picture. Once the training is complete, the application will be ready for recognition of sign language input. The input is taken via a webcam. This is a continuous video input, each frame of which OpenCV treats as an individual image and returns string output based on previously trained gestures in the dataset.

➢ *Speech*

To enable a sign language speaker to talk to an audience, a speech synthesis mechanism would enhance the effectiveness of the conversation. For this, we need a text-to-speech engine. There many popular test-to-speech libraries available such as Google's "gtts" (Google text-to-speech), pysttx3, and multiple web speech APIs. For this project, we have used pyttsx3 library which helped us achieve real-time speech synthesis.

IV. RESULTS

The training accuracy we obtained 99.93% and the data loss was 0.339.

The following images are the outputs of the program:

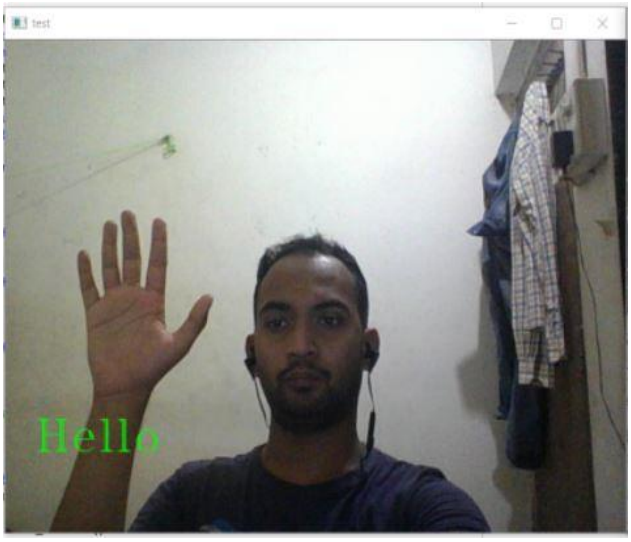

Fig 2

The above image shows the prediction of the given gesture input which was set by the user. The input frame is converted to the HSV form and the hsv values can be set by the user according to his/her background condition.

Fig 3

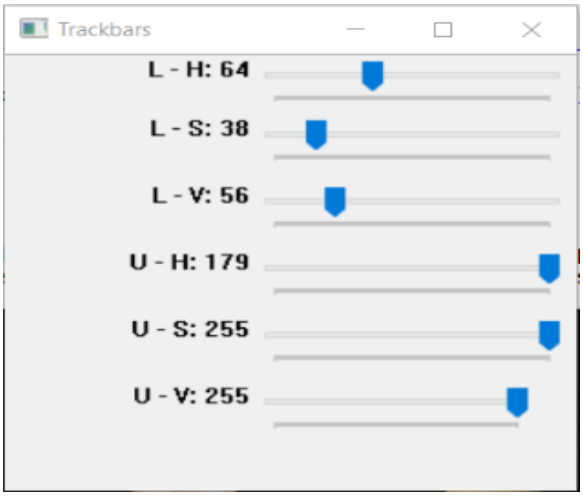

Fig 4

To set the raange of the hsv values the user just have to slide the bars and check the effectiveness of the corresponding values. According to the given lower limits and upper limits of the HSV values, the background noise will be reduced.

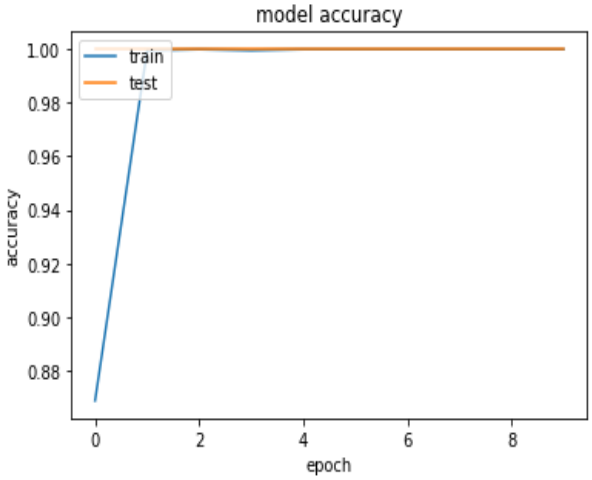

Fig 5

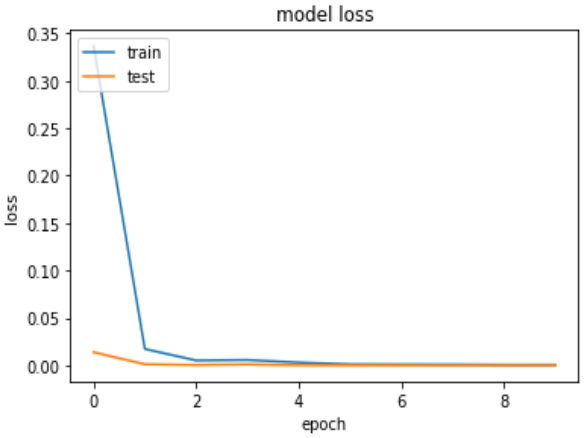

Fig 6

Above is the graph of the model accuracy and model loss while training the CNN model.

V. CONCLUSION

The application of this concept can empower and give a voice to the sign language community, they can convey their thoughts to the normal people and lack of human translator will not be felt.

REFERENCES

[1]. Suharjitoa, Ricky Anderson , Fanny Wiryana , Meita Chandra Ariesta , Gede Putra Kusumaa, Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output, 2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI 2017, Bali, Indonesia.
[2]. Manisha U. Kakde, Mahender G. Nakrani, Amit M Rawate, "A Review Paper on Sign Language Recognition For Deaf And Dumb People using Image Processing," ,IJERT, ISSN:2278-0181, Vol.5 Issue 03, March 2016
[3]. Pratibha Pandey, Vinay Jain, "Hand Gesture Recognition for Sign Language: A review,", IJSETR, Vol.4, Issue 3, March 2015.
[4]. Shreyasi Narayan Sawant, M. S. Kumbhar, "Real Time SignLanguage Recognition using PCA", 2014 IEEE conference on Advanced Communication Control and Computing Technologies.
[5]. Karishma Dixit, Anand Singh Jalal,"Automatic Indian Sign Language Recognition System", IEEE 2012.
[6]. Jie Huang, Wengang Zhou, Houqiang Li, and Weiping li,sign language recognition using real-sense
[7]. Wen Gao, Gaolin Fang, Debin Zhao, and Yiqiang Chen,"A chinese sign language recognition system based on sofm/srn/hmm," Pattern Recognition, vol. 37, no. 12, pp. 2389–2402, 2004.