REAL TIME ASSISTIVE INTERPRETER FOR DEAF COMMUNITY OVER MACHINE LEARNING

Ms. Priyanka Pandurang Patil
Department of Computer Engineering
Zeal college of engineering and research, Pune

Dr. Y. B. Gurav
Department of IT Engineering,
Zeal College of engineering and research, Pune.

ABSTRACT

Sign Language research field based on real time hand gestures called sign samples and recognition unit of computer. The sign language is the only way to used for deaf and dumb community communication platform. In this system, we are working on the American Sign Language (ASL) dataset (A-Z) alphabet recognition followed by our word recognition dataset of Indian Sign Language (ISL). Sign data samples to be making our system more accurate with help of Convolutional Neural Network (CNN). Today, much research has been going on the field of sign language recognition but existing study failed to develop trust full communication interpreter. The purpose of this system is to represent a real time two way communication interpreter based on Indian Sign Language (ISL) with higher accuracy. Indian Sign Language (ISL) used by Deaf peoples community in India, does have acceptable, meaningful essential and structural properties.

Keywords: Artificial Neural Network, Indian Sign Language, Hand Gesture Recognition, Deaf community

I. INTRODUCTION

There are so many languages in India used as officially and locally. Such large diversity country has more challenges to maintain uniqueness in language interpretation. In languages have its challenges when it used to communicating over different areas, societies and states. Indian Sign Language (ISL) is one of the living languages in India used by the Deaf community peoples. But as we seen there is not any standard language available till date. So we are working on different sign language dataset to invent Indian sign language as a interpreter.

We are going to implement two way communications for Sign language is used for the people who are deaf or hard of hearing and also used by them who can hear but cannot physically speak. Our motive behind this implementation is to create complete language which involves movement of hands, facial expressions and gesture of the body. The Sign language is not universal standard so we are making our contribution towards sign language development. Every country has its own native sign language like American Sign Language work for alphabet recognizer. Each sign language has its own rule and semantic meanings. The problem comes when deaf and dumb people want to communicate or trying to say something there is not any language for them. So it becomes necessary to develop an automatic language interpreter to assist them for their fluent communication. They people want something more helpful which makes there communication universal and easy. Another one is based on computer vision based gesture recognition, which involves image processing techniques. Consequently, this category faces more complexity. Our motive to develop this system based on real time signs.

This system captures hand gesture images of ISL with system camera for feature extraction. The analyzing phase, pre-processing unit is used to the noise removal, grey scale conversion by using Gaussian filter, binary conversion of images done by using OTSU's method followed by feature extraction. In our system, Convolutional Neural Network (CNN) is going to used for future recognition in which we having the input unit of training data set of images. Next we have hidden unit which acts upon this training dataset to evaluate the output unit results train model. This entire CNN works by considering the factors namely matrix feature of images for drafting into a train model for real time sign recognition. The working with real time sign language we know that the dataset need to be large and rich in processed features.

II. LITERATURE SURVEY

Sharmila Konwar et.al [1] states that is used to design an automatic vision based American Sign Language detection system and converting results in to text. The work introduced in this paper is meant to outline a programmed vision based American Sign Language recognition framework and interpretation to content. To distinguish the human skin shading from the picture, HSV shading model is utilized. At that point edge recognition is connected to distinguish the hand shape from the picture. An arrangement of morphological activity is connected to get a refined yield for the gesture based communication acknowledgment This work is mainly focused on the color model and edge detection phenomenon. Edge detection algorithm the hand gestures are detected successfully for the alphabets in American language. Some images are not detected successfully due to geometric variations, odd background and light conditions.

Yo-Jen Tu et.al [2] presents a face and signal acknowledgment based human-PC communication (HCI) framework utilizing a solitary camcorder. Not the same as the traditional specialized strategies among clients and machines, we consolidate head posture and hand motion to control the hardware. We can recognize the situation of the eyes and mouth, and utilize the facial focus to assess the posture of the head. Two new techniques are displayed in this paper: programmed signal territory division what's more, introduction standardization of the hand signal. It isn't compulsory for the client to keep signals in upright position, the framework fragments and standardizes the signals consequently. They explore demonstrates this technique is extremely precise with motion acknowledgment rate of 93.6%. The client can control different gadgets, counting robots all the while through a remote system.

Angur M. Jarman et.al exhibits another calculation to distinguish Bengali Sign Language (BdSL) for perceiving 46 hand signals, including 9 motions for 11 vowels, 28 motions for 39 consonants and 9 motions for 9 numerals as indicated by the similitude of elocution. The picture was first re-sized and after that changed over to double configuration to edit the locale of enthusiasm by utilizing just best most, left-most and right-most white pixels. The places of the fingertips were found by applying a fingertip discoverer calculation. Eleven highlights were extricated from each picture to prepare a multilayered feed-forward neural system with a back-spread preparing calculation. Separation between the centroid of the hand area and each fingertip was ascertained alongside the points between every fingertip and flat x pivot that crossed the centroid. A database of 2300 pictures of Bengali signs was developed to assess the viability of the proposed framework, where 70%, 15% and 15% pictures were utilized for preparing, testing, and approving, separately. Exploratory outcome demonstrated a normal of 88.69% exactness in perceiving BdSL which is particularly encouraging contrast with other existing techniques.

Javeria Farooq et. al states that hand motion acknowledgment is a characteristic and natural way to connect with the PC, since cooperation's with the PC can be expanded through multidimensional utilization of hand motions as contrast with other information techniques. The reason for this paper is to investigate three unique strategies for HGR (hand signal acknowledgment) utilizing fingertips location. Another methodology called "Arch of Perimeter" is given its application as a virtual mouse. The framework exhibited, utilizes just a webcam and calculations which are created utilizing PC vision, picture and the video handling tool stash of Matlab.

Guillaume Plouffe et. al examines the advancement of a whiz signal UI that tracks and perceives progressively hand signals in light of profundity information gathered by a Kinect sensor. The intrigue space relating to the hands is first portioned based on the suspicion that the hand of the client is the nearest protest in the scene to the camera. A novel calculation is proposed to move forward the checking time with a specific end goal to recognize the main pixel on the hand form inside this space. Beginning from this pixel, a directional scan calculation takes into account the recognizable proof of the whole hand form. The k-arch calculation is then utilized to find the fingertips over the form, and dynamic time twisting is used to choose motion competitors and furthermore to perceive motions by contrasting a watched motion and a progression of pre-recorded reference motions. The examination of results with cutting edge approaches demonstrates that the proposed framework beats a large portion of the answers for the static acknowledgment of sign digits and is comparable regarding execution for the static and dynamic acknowledgment of well-known signs and for the communication through signing letter set. The arrangement at the same time manages static and dynamic motions also similarly as with various hands inside the intrigue space. A normal acknowledgment

rate of 92.4% is accomplished more than 55 static and dynamic signals. Two conceivable utilizations of this work are talked about furthermore, assessed: one for elucidation of sign digits and signals for friendlier human– machine cooperation and the other one for the normal control of a product interface.

ZAFAR AHMED ANSARI et. al states individuals with discourse inabilities convey in gesture based communication and accordingly experience difficulty in blending with the healthy. There is a requirement for a translation framework which could go about as a scaffold among them and the individuals who don't have the foggiest idea about their gesture based communication. An utilitarian unpretentious Indian gesture based communication acknowledgment framework was executed and tried on true information. A vocabulary of 140 images was gathered utilizing 18 subjects, totaling 5041 pictures. The vocabulary comprised for the most part of two-gave signs which were drawn from a wide collection of expressions of specialized and every day utilize starting points. The framework was executed utilizing Microsoft Kinect which empowers encompassing light conditions and question shading to have irrelevant impact on the effectiveness of the framework. The framework proposes a technique for a novel, minimal effort and simple to-utilize application, for Indian Sign Language acknowledgment, utilizing the Microsoft Kinect camera.

Keerthi S Warrier et. al states Hand Gesture Recognition System (HGRS) for detection of American Sign Language (ASL) has become essential and powerful communication tool for specific users (i.e. hearing and speech impaired) to interact with general users via computer system. Numerous HGRS have been developed for identification of diversified sign languages using effective techniques. There exist two main approaches in the hand gesture analysis namely; vision-based and device-based approach. In vision-based approach the user does not require to wear any extraneous mechanism on hand. Instead the system requires only camera(s), which are used to capture the images of hand gesture symbol for interaction between human and computers.

III. METHODOLOGY USED IN EXISTING SYSTEM

There are different approaches for image capturing is being used before. The captured images from either real time cam or static from dataset is used for further processing. Digital image processing is a field that analyses image processing methods. The image used in this is a static image form computer vision (webcam). Mathematically, the image is a formulation of light intensity on two-dimensional field. The image to be processed by a system or computer, an image should be presented statistically with numerical values. A digital image can be stated by a two-dimensional matrix f (m, n) consisting of M columns and N rows. The color image processing [RGB], there are different models are like hue and saturation, value (HSV) model. This model is used with an object with a certain color can be identified and to reduce the unwanted light intensity from the outside. Further Tests on images were performed using six kinds of colors, ie brown, yellow, green, blue, black and white.

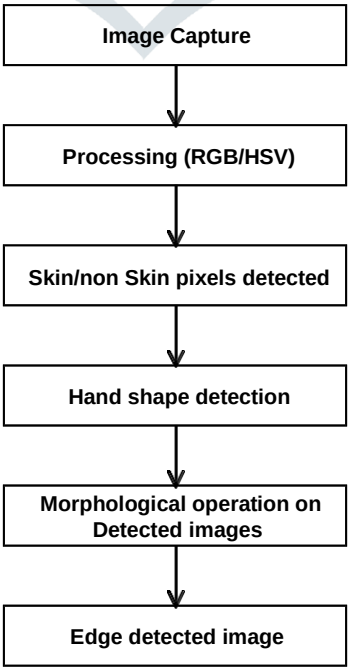


Fig. Existing System Architecture

A skin detector mainly used to transforms a given pixel into an appropriate color space. Then it is use a skin classifier to label the pixel. It correctly differentiates a skin or non-skin pixels. A skin classifier decides a decision boundary of the skin color class in the color space based on a training database of skin-colors pixels. Hand shape detection done by using object recognition feature of edge drawing methods. The most basic morphological operations are used to adds pixels to the boundaries of hand shape in an image and removes pixels on hand shape object boundaries. The number of pixels added or removed from the hand shape in an image depends on the size and shape of real time sign image. After studying multiple papers we conclude that those all work is unable to create communication interpreter while they are suffering or focusing only image processing and object recognition.

IV. METHODOLOGY USED IN PROPOSED SYSTEM

We are going to overcome existing communication barrier by providing two way communications for deaf and dumb peoples. We take input as action of hand gestures and convert it into common words of communication after getting this in text, convert it into voice. After getting voice, normal people can understand it. Similarly, normal people can speak in voice, our system will convert it into text and further convert it into actions which is simply understandable by deaf people.

In this system Sign language is the primary language of the people who are deaf or hard of hearing and also used by them who can hear but cannot physically speak. It is a complex but complete language which involves movement of hands, facial expressions and postures of the body. Sign language is not universal. Every country has its own native sign language. Each sign language has its own rule of grammar, word orders and pronunciation. The problem arises when deaf and dumb people try to communicate using this language with the people who are unaware of this language grammar. So it becomes necessary to develop an automatic and interactive interpreter to understand them.

So its mandatory to overcome these communication gap between the deaf community and normal persons. Two way communications system is providing for deaf and dumb peoples. We take input as action of hand gestures and convert it into common words of communication after getting this in text, convert it into voice. After getting voice, normal people can understand it. Similarly, normal people can speak in form of voice; our system will convert it into text and further convert it into actions which are simply understandable by deaf people.

We are going to develop two way communication systems by using machine learning and image processing techniques. The current real time application will work for real time assistance.

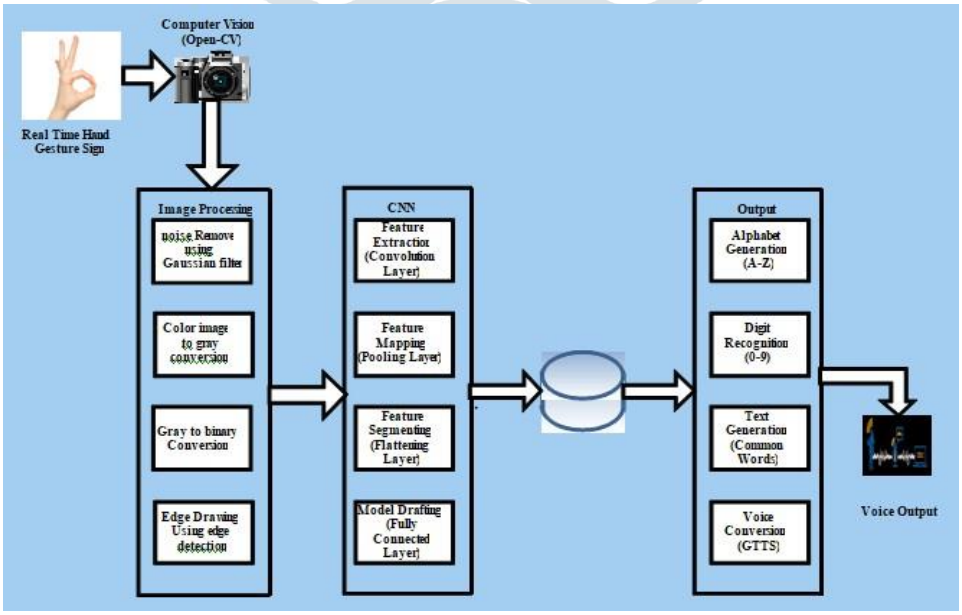


Fig. System Architecture

We are going to develop following modules:

1. **Hand action recognition**
   Open-CV (Open Source Computer Vision) is a library of programming functions used for real time image processing with computer vision. In our implementation we are going to use open compute vision for taking real time snap of hand gestures for further processing. After getting real time hand gestures image processing applied on it for removing noise from it.

2. **Image processing**
   After getting real time image of hand gesture send for image processing module. In image processing image gets converted in gray format by removing noise in it using Gaussian filter. After gray conversion image thresholding by setting RGB color values to zero and preserving only black and white [0 and 1] values. Gray to binary conversion is done by using OTSU's method. After getting black formatted image hand shape get extracted from image. The exact shape of hand will get by drawing edge using canny edge detection method.

3. **Feature extraction**
   After getting exact shape of hand gestures features get extracted from it by using pixels weight calculations. The image pixels get drafted in matrix by using weight gradient functions. Feature extraction done on all hand gesture dataset for training model creation and drafting. The train model creation done by using deep learning (CNN) algorithm.

4. **Feature mapping & text generation**
   In real time image of hand gestures is going through image processing and subsequent phases of feature extraction. After getting image features these statistical features get matched with pre trained model and respective text generated. After text generation those text get converted into a voice.

5. **Voice conversion**
   The text generated further gets converted into voice by using Google's text to speech library. After voice generation will be used for communication purpose deaf person to normal person.

6. **Text to action conversion**
   For normal person to deaf communication normal person use their own language in the form of voice. The voice generated from normal people gets recognized by speech recognizer and this speech gets converted into a text. After text from normal person get semantically mapped with sign samples. The matched sign samples will show by using open-CV automatically. The sign images get easily understand by deaf and dumb community persons.

V. RESULTS AND DISCUSSION

In two way communication system we have been implemented greatly trained model that can accurately analyses hand gesture signs. In this system we used tensor-flow machine learning framework and predefined libraries.

A. Gray scale conversion

In gray scale conversion color image is converted into a gray form using Gaussian blur. Color image containing noise and unwanted background which is removed or blurred by using this method.

Fig.1 Gray Image

B. Binary conversion

Gray scale image is given to input for Otsu's method for binary conversion. In Binary form of images converted in 0 and 1 form means black and white.

Fig.2 Binary Image

C. Edge Detection

In Edge detection binary image get dimensions by counters using convex hull algorithm. In which eccentricity finding drawing edges around white portion of binary image.

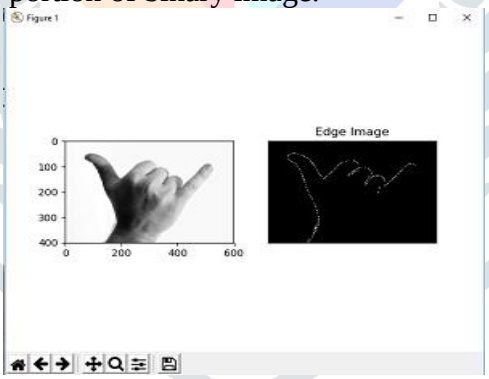


Fig.3 Edge Detection Image

D. Training Model

In two way communication system we are using tensor flow for training and validating our models dataset. In which 87000 image samples are trained for alphabet recognition model after that we have used 7500 words image samples for word recognition model. Finally plot files generated as an output of our trained model.

E. Testing Model

In final phase of data testing in which real time hand gesture images matched by our training model with higher percent of accuracy.

After matching hand gestures respective alphabets display on console and stored in text file as well. Finally we have been used Google text to speech for converting into a voice.

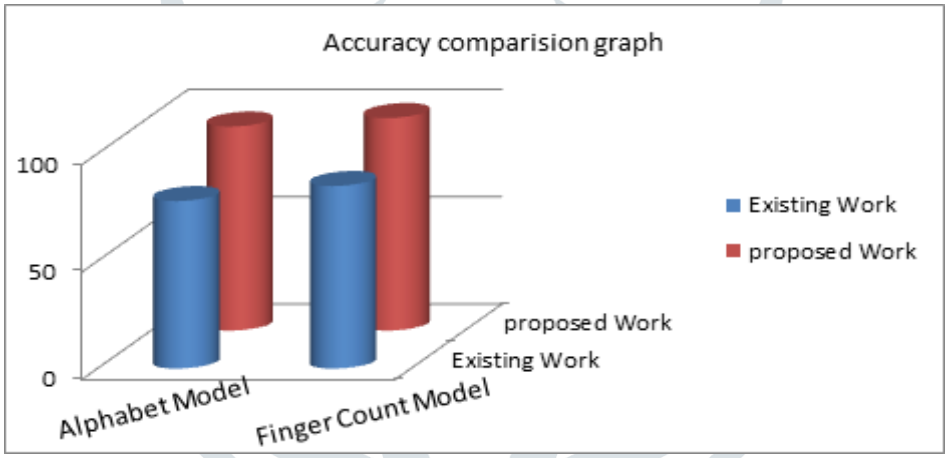
Second way we get real-time voice input from user to get desired text from it. After getting text from voice then output sign sample displayed by our system for visually understand by deaf person. As well as we have suggested video samples for general activity for better understanding of deaf persons.

In our experimental setup, In table 1 describe our system modules and respective generated output.

Sr. No.	No Input Sample's	Output Generated
1	1 to 5 Hand gesture Images	Fingertip counts 1 to 5
2	29 Hand gesture's	A-Z , Nothing , Space and Delete
3	15 Basic Words	Voice
4	Real time voice input	Text
5	Text input	Sign/video

F. Comparative Study Graph

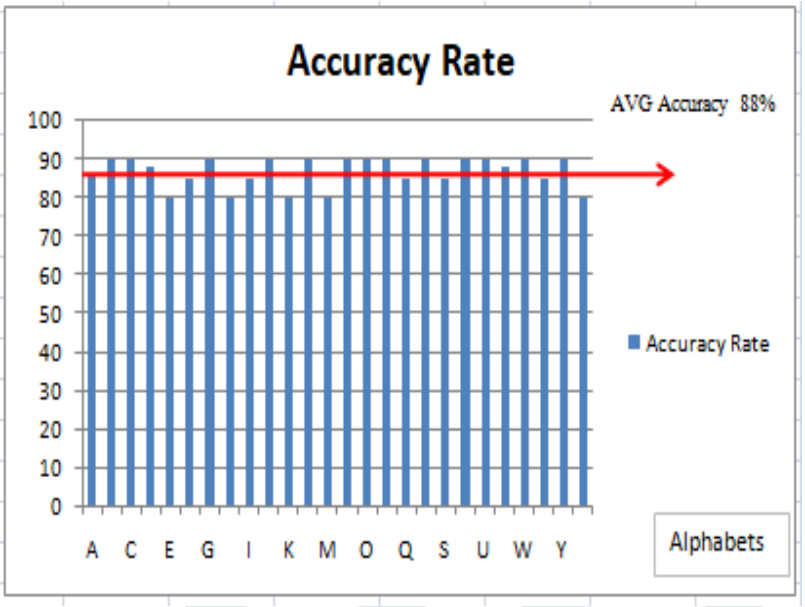
Table 1: Modules of System As per compared with existing work carried out for deaf community peoples gets alphabet recognition model and finger count model. The existing work gets accuracy in alphabet model we get accuracy nearly 80 percent and in proposed model we get accuracy nearly 95 percent. The finger count model we get accuracy in existing model 85 percent and proposed model gets 99 percent accuracy.



Graph 1: Accuracy Comparison Graph

G. Accuracy Rate of Sign Recognition

All sign sample images trained by our trained model approximately 3000 images per alphabet. Total around 87000 images trained so we have been conclude the accuracy rate average 90%. All common words sign samples trained in our system is collected nearly about 500 images per folder. We achieved 95% accuracy in word detection model. We are using real time speech recognizer library for speech recognition and convert speech to text.



Graph 2: Sign sample Average Accuracy Rate

VI. CONCLUSION

Thus we implemented the two way communication interpreter using American Sign Language to accurately recognize the real time hand gestures and convert those alphabets and words. The recognized sign converted into text and further it converted into voice. Convolutional Neural Network algorithm is applied to achieve greater accuracy in real time sign recognition. In real time scenario we have achieved accuracy above 90% in each trained models.

In future work we will more focus on real time action recognition to overcome hand sign limitations.

VII. REFERENCES

[1] Sharmila Konwar, Sagarika Borah and Dr. T. Tuithung, "An American sign language detection system using HSV color model and edge detection", International Conference on Communication and Signal Processing, IEEE, April 3-5, 2014, India

[2] Yo-Jen Tu, Chung-Chieh Kao, Huei-Yung Lin,"Human Computer Interaction Using Face and Gesture Recognition", Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific, IEEE, Kaohsiung.

[3] Angur M. Jarman, Samiul Arshad, Nashid Alam and Mohammed J.Islam, "An automated Bengali sign language recognition based on finger tip finder Algorithm", IJEI, 2015.

[4] Javeria Farooq and Muhaddisa Barat Ali, "Real time hand gesture recognition for computer interaction", International Conference on Robotics and Emerging Allied Technologies in Engineering (ICREATE), 22-24 April ,2014.

[5] Guillaume Plouffe and Ana-Maria Cretu, "Static and dynamic hand gesture recognition system in depth data using dynamic time warping" IEEE Transactions on Instrumentation and Measurement (Volume: 65, Issue: 2, Feb. 2016)

[6] ZAFAR AHMED ANSARI, "Nearest neighbour classification of Indian sign language gestures using kinect camera" ,

February 201 , pp. 161–182

[7] Keerthi S Warrier , "Software based sign language converter" , April 2016 IEEE.

[8] Vaishali.S.Kulkarni et al. , "Appearance Based Recognition of American Sign Language Using Gesture Segmentation" , (IJCSE) Vol. 02, No. 03, 2010, 560-565

[9] Sabaheta , Gunay Karli , " Sign Language Recognition using Neural Networks" TEM Journal – Volume 3 / Number 4 / 2014.

[10] RafiqulZaman Khan, Noor Adnan Ibraheem , "Survey on Various Gesture Recognition Technologies and Techniques"