International Journal of Current Engineering and Technology
©2021 INPRESSCO®, All Rights Reserved

E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
Available at http://inpressco.com/category/ijcet

Research Article

Real Time Assistive Interpreter for Deaf Community Over Machine Learning

Ms.Priyanka P.Patil and Dr.Y.B.Gurav

Computer Engineering Department Zeal College of engineering and research Pune,India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

After studying multiple research papers on machine learning and real time hand gesture recognition many systems worked only on sign recognition for number generation and sign recognition for A-Z alphabet generation modules. By considering communication interface for deaf and dumb community one way communication is not enough. So there is need of two way communication interface to overcome communication gap. Existing module is failed to fill this barrier of communication. We are going to overcome existing communication barrier by providing two way communications for deaf and dumb peoples. The current work is based on American Sign Language dataset of letters A-Z and 0-9. The propose work is going to invent word dataset for common words and making interpreter for communication with the help of Convolutional neural network (CNN). We are going to focus on accuracy factor and time complexity.

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

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. In this system, 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.

Literature Survey

Javeria Farooq and Muhaddisa Barat Ali [1] state that 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 prerecorded 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 wellknown 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.

Guillaume Plouffe and Ana-Maria Cretu [2] proposed that 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. A 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, totalling 5041 pictures. The vocabulary comprised for the most part of twogave 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.

Zafar Ahmed Ansari [3] introducing 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.

Keerthi S Warrier [4] 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.

Sharmila Konwar et.al [5] proposed 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.

Yo-Jen Tu et.al[6] introducing 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 feedforward 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.

Angur M. Jarman et.al[7] 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.

Comparison of Proposed Work

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.

Proposed Methodology

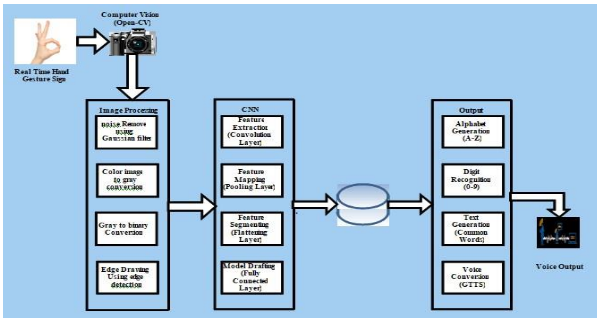

Computer Vision (Open-CV)
Real Time Hand Gesture Sign
Image Processing
noise Remove using Gaussian filter
Color image to gray conversion
Gray to binary Conversion
Edge Drawing Using edge detection
CNN
Feature Extraction (Convolution Layer)
Feature Mapping (Pooling Layer)
Feature Segmenting (Flattening Layer)
Model Drafting (Fully Connected Layer)
Output
Alphabet Generation (A-Z)
Digit Recognition (0-9)
Text Generation (Common Words)
Voice Conversion (GTTS)
Voice Output

Figure No.1 System Architecture

In this proposed system 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 are simply understandable by deaf people. This system can develop by using following modules:

A. 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.

B. 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.

C. 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.

D. 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.

E. 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.

F. 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.

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.

Convolutional Neural Network CNN takes in processed images as input.

- Extracts different features about the images regardless of their position using a series of mathematical operations to identify the pattern.
- Every layer in CNN has API which transforms input to output with differentiable functions.

- Figure No.2 Block diagram CNN

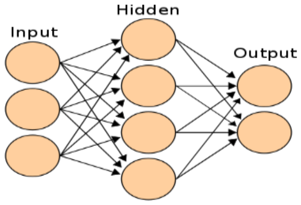


Figure No.2 Block diagram CNN

In CNN consist following 4 steps as shown in figure no.3

- Convolutional Layer
- Pooling
- Flattening
- Fully Connection

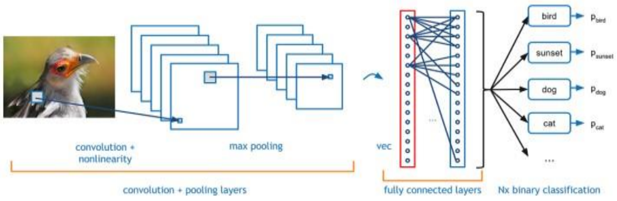


Figure No.3 Steps of CNN

1. Convolutional Layer

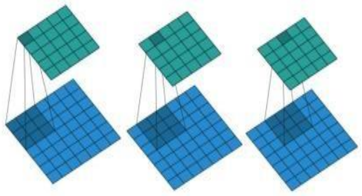


Figure No.4 Convolutional Layer

In above figure no.4 extract different features pixel wise by using feature detectors/kernels. Perform numerous convolutions on input, where each operation uses a different filter. This results in different feature maps. In the end, we take all of these feature maps and put them together as the final output of the convolution layer.

2. Pooling

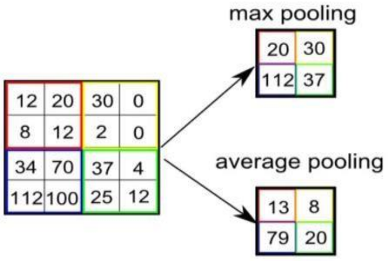


Figure No.5 Pooling

In above figure no. 5 show that the function of pooling is to continuously reduce the dimensionality to reduce the number of parameters and computation in the network. This shortens the training time and controls over fitting Max Pooling extracts out the highest pixel value out of a feature while average pooling calculates the average pixel value that has to be extracted.

3. Flattening

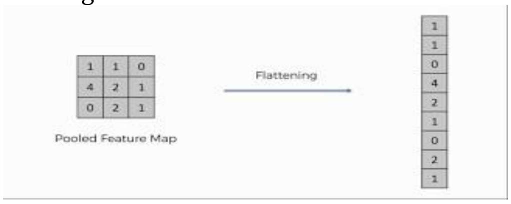


Figure No.6 Flattening

In above figure no. 6 shows that basically here we arrange the pooled feature into a single vector/column as a input for next layer (convert our 3D data to 1D)

4. Fully Connection

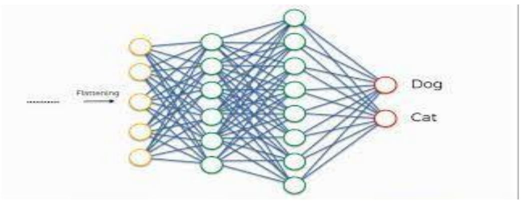


Figure No.7 Fully Connection

In above figure no. 7 shows Neurons in a fully connected layer have full connections to all the activations in the previous layer. Combining more neurons to predict more accurately.

Convolutional Neural Network Approach (CNN)
Step 1- Input hand gesture's
Step 2- Image capturing by using open-cv
Step 3- Image processing
Step 4- Feature Extraction from images
Step 5- Model generation
Step 6- Sign recognition
Step 7- text generation
Step 8- voice conversion by using GTTs

H. Mathematical Model • System Description:

S={s, e, X, Y,Φ} Where, s = Start of the system. 1. Log in System
2. Image processing 3. Text and voice Generation e = End of the program X = Input sign samples
Y = Output of the program (voice & text).
Proper detection of sign samples and generates voice and text accordingly. X, Y ∈ U
This module helps the deaf community as an interpreter.
Let U be the Set of System.
U= {Sk1, Sk2, SK3, R1, R2, A}
Where Sk1, Sk2, SK3, R1, R2 are the elements of the set.

- SK1: Real time sign images. SK2: voice as a input
- SK3: voice and image processing
- R1: Sign to text and voice conversion
- R2: voice to text and text to sign conversion
- A=text and voice Generation

SPACE COMPLEXITY:
No need to store Indian sign language data-set in the system, so space complexity is less.
TIME COMPLEXITY:
We are going to use CNN for fast and better recognition with higher accuracy. So time complexity is less. So the time complexity of this algorithm is $O(n^n)$.

Success:

1. We get higher accuracy over existing work by providing real time sign recognition system.

2. We get accurate sign recognition with high reliability. Failures:

1. Large sign image data set take more time to extract features makes system slow.

2. Images with lower resolution reduce accuracy.

Mathematical Model in Equation format
Notation
Where,

M= Set of all entities.
• U= {U1,U2,U3,....UN} where (U1,U2,..UN) are the number of users.
• N={N1,N2,N3,....N10} where (N1,N2, ...N10)are the hand gestures for 0-9 numbers
• A={A1,A2,A3,...AN} where(A1, A2,....AN) are the hand gestures for A-Z alphabets.
• S={S1,S2,S3,...SN} where (S1,S2,S3,...SN) are the real time sign inputs.
• T={T1,T2,T3,....TN}where T is the Final text formed by common words (T1,T2,T3,....TN) • V={V1,V2,V3,....VN} where (V1,V2,V3,...VN) are the voice generated for real time sign inputs • TNU= Total number of users
• THGN= Total number of hand gesture number • TNA=Total number of alphabets
• TNRTS=Total number of Real time sign inputs
• TNFTF=Total number of final text formed by common word
• TNVG=Total number of voice generated

For calculate total number of users by following equation 1 Total number of users = Number of user1+ Number of user 2 +.....+ Number of N users ∑ TNU = ∑ U1+ ∑ U2+...+∑ UN......equation 1 For calculate total number of hand gesture number Total number of hand gesture number = Number of hand gesture number1 + Number of hand gesture number 2 +.....+Number of hand gesture number N ∑ THGN= ∑ N1+ ∑ N2+...+∑NN............equation 2 For calculate total number of alphabets Total number of alphabets = Number of alphabets1+ Number of alphabets2+.....+Number of alphabets N ∑ TNA= ∑ A1+ ∑ A2+...+∑AN............equation 3 For calculate total number of real time sign inputs Total number of real time sign inputs = Number of real time sign input 1 + Number of real time sign input 2+…..+ Number of real time sign inputs N ∑ TNRTS = ∑ S1+ ∑ S2+…+∑SN…………equation 4 For calculate total number of final text formed by common words Total number of final text formed by common words = Number of final text formed by common words 1+Number of final text formed by common words 2+…..+ Number of final text formed by common words N ∑ TNFTF = ∑ T1+ ∑ T2+…+∑TN…………equation 5 For calculate total number of voice generated Total number of voice generated= Number of voice generated 1 + Number of voice generated 2+…..+ Number of voice generated N ∑ TNVG = ∑ V1+ ∑ V2+…+∑VN…………equation

Result and Discussions

In two way communication system we have been implemented greatly trained model that can accurately analyze 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.

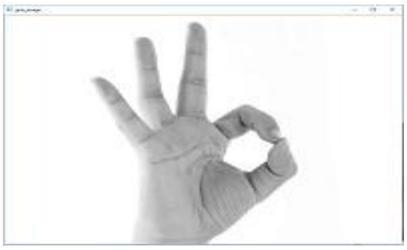


Fig.8 Gray Image

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.

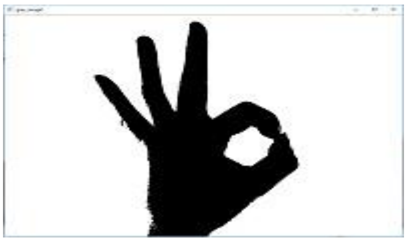


Fig.9 Binary Image

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.

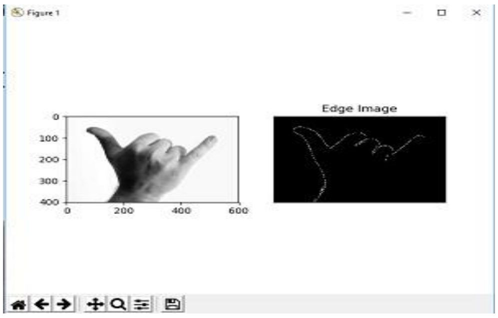
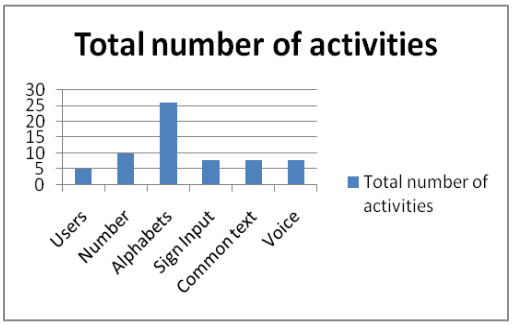


Fig.10 Edge Detection Image

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 no.11.1, shows total number of users, total number of hand gesture number, total number of alphabets, total number of Real time sign inputs, total number of final text formed by common word, total number of voice generated. In this table consist 5 users, 10 numbers of hand gesture number, 26 alphabets, 8 real time sign inputs,8 number of final text formed by common word and 8 number of voice generated.

Sr.No	Activity	Total number of Activity
1	Total number of users	5
2	Total number of hand gesture number	10
3	Total number of alphabets	26
4	Total number of Real time sign inputs	8
5	Total number of final text formed by common word	8
6	Total number of voice generated	8

From above data, as shown in graph 12.1,the total numbers of users were 5, total numbers of hand gesture number were 10, total number of alphabets were 26,total number of real time sign inputs were 8,total number of final text formed by common word were 8 and total number of voice generated were 8.



Graph 12.1: Total number of activities VI.

Conclusions

In Sign Language Recognition system, we will are going to tackle the communication problem of deaf and dumb community by inventing two way communication interpreters. We are going to propose hand gesture recognition systems based on American Sign Language dataset with our Indian sign contribution using deep learning approach. System will be two way communication systems by using sign to text and voice to sign conversion phenomenon. Our future research will be extended for further improvement in recognition accuracy and also for motion detection of body for word recognition.

Acknowledgment

This work is supported in a Sign Language research field in India. Authors are thankful to Faculty of Engineering and Technology (FET), Savitribai Phule Pune University, Pune for providing the facility to carry out the research work.

References

[1] 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.

[2] 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)

[3] Zafar Ahmed Ansari, “Nearest neighbor classification of Indian sign language gestures using kinect camera”, February 2016, pp. 161–182

[4] Keerthi S Warrier, JyateenKumar Sahu, Himadri Halder, Rajkumar Koradiya, and Karthik Raj V“Software based sign language converter” , April 2016 IEEE.

[5] 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

[6] 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.

[7] Angur M. Jarman, Samiul Arshad, Nashid Alam and Mohammed J.Islam, "An automated Bengali sign language recognition based on finger tip finder Algorithm", International journal of Electronics & Informatics, 2015.