Real time Indian sign language system using Artificial Intelligence

[1]Jatin Jagani, [2]Yogita Parikh, [3]Vishal Savaliya, [4]Gaudani Mohit
[2] Assistant professor at Biomedical Engineering Department,
[1,2,3,4]Biomedical Engineering Department,
[1]L. D. College of Engineering, Ahmedabad, India.

Abstract : ***Abstract***: This paper presents approach to convert "Sign Language into speech using Artificial Intelligence in real time" for deaf and dumb people. Deaf and Dumb People feels difficulty to communicate with the ordinary people. The sign language of deaf and dumb is quite difficult to learn and it is not possible for everybody to learn that language. This approach helps to provide communication between people with speech impairment and normal people, thereby reducing the communication gap between them. This system capable of recognizing 26gestures of letter from the Indian Sign Language by using Artificial Intelligence. The proposed system having Real time video acquisition, frame extraction and Pre-processing, Hand gesture segmentation, Sign recognition using AI, Text and audio output of letter. By using Webcam, capture the real time video of hand gestures for Indian sign language and extract the video frames from it. Brightness and orientation of frame is derived and if its value is not proper, then frame is rejected and it is acquired again. Sign gesture is segmented by using K-means clustering algorithm and morphological image processing. Sometimes there is difficult to segment the hand gestures due to different background conditions. A convolutional Neural Network recognize the letter of ISL with the 97% recognition rate. It is trained by using different segmented hand gesture of letters in different background conditions. Recognized gesture is used to convert sign into text and audio output for users.

IndexTerms - Sign language converter, Artificial Intelligence, Text to speech conversion, Indian Sign Language**.**

I. INTRODUCTION

II. Sign language is used across the world to help bridge the communication gap for the hearing or speech impaired. It is a system of communication using visual gestures and signs, as used by deaf and dumb people. There are various categories in the sign language like ISL (Indian Sign Language), ASL (American Sign Language), BSL (British Sign Language) and etc... But none of the sign languages are universal or international. A person should know the sign language to understand the language; this becomes complicated when a person who has inability to speak or hear wants to convey something to a person or group of persons, since most of them are not familiar with the sign language [1]. In the last several years there has been an increased interest among the researchers in the field of sign language recognition. There are lots of techniques and modulations are being introduced and are under research to minimize or simplify the complexity in sign language to speech conversion. The approach has been proposed in the aim of minimizing all those complexity and to attain maximum accuracy in conversion of Indian sign language to speech with hand gestures by using Artificial Intelligence. Human gestures are an important sign of human communication and an attribute of human actions informally known as the body language. A lot of methods are being in use to track human gestures. To get maximum accuracy and to bring out the system unique a lot of methods are attempted and best case is user defined actions (gestures) to control the system [2].

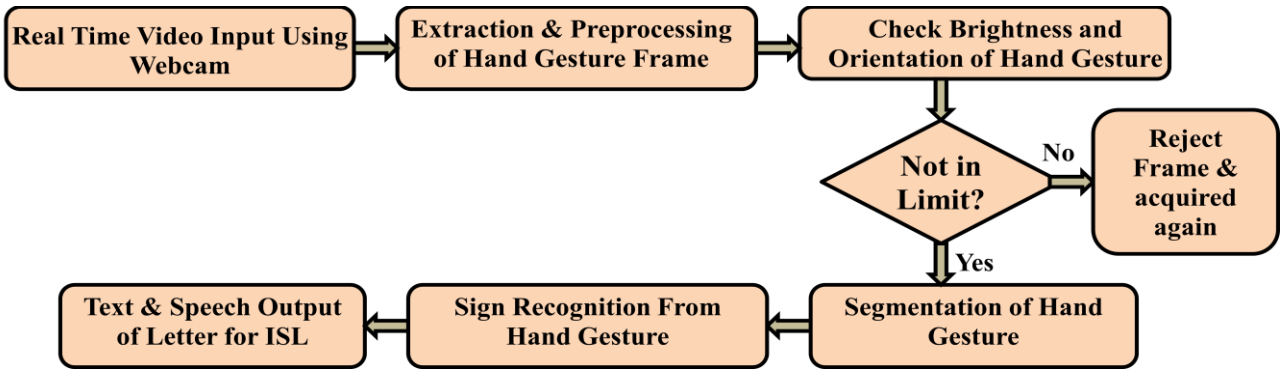


Fig.1.Block diagram of Proposed Indian Sign language system

In image processing based sign language system, the real challenge is correct recognition of hand gestures with real time complex background conditions. The proposed sign language system is able to recognize accurately hand gestures for different letters by using a webcam which is interfaced with MATLAB in least time and converts alphabets of Indian Sign Language into corresponding text as well as into audio output in a vision based system. The architecture of the system has shown in Fig.1. In this proposed work, The real time hand gestures of Letter for ISL is acquired using laptop webcam. The frame is extracted and after checking its brightness and orientation, send to further processing. Hand gesture is segmented using K-means clustering algorithm based on color of skin in the frame and morphological image processing. From the train data of segmented hand gestures for letters, train the convolutional neural network to recognize the pattern of different hand gestures of sign. Using convolutional neural network, the imposter hand gestures is tested correctly.

III. LITERATURE REVIEW

In the sign language system has two well-known approaches are based on Image processing and Data glove[7]. In data glove approach, user need to wear glove consist of flex sensor and motion tracker. Data are directly obtained from each sensor depends upon finger flexures and computer analysis sensor data with static data to produce sentences. It's using neural network to improve the performance of the system. The main advantage of this approach less computational time and fast response in real time applications [3][4].Another approach using a portable Accelerometer (ACC) and Surface Electro Myogram (sEMG) sensors used to measure the hand gesture [5]. ACC used to capture movement information of hand and Arms. EMG sensor placed, it generates different sign gesture. Sensor output signals are fed to the computer process to recognize the hand gesture and produce speech/text. But none of the above methods provide users with natural interaction. Al-Ahdal and Tahir [6] presented a novel method for designing SLR system based on EMG sensors with a data glove. This method is based on electromyography signals recorded from hands muscles for allocating word boundaries for streams of words in continuous SLR.

In image processing based technique, Video of hand gestures is captured through web camera. Hand gestures image is extracted from video by using frame extraction method. Hand gestures of letters are detected by using various image processing techniques like image filtering, K-means and Edged based segmentation, Morphological image processing, Feature extraction methods. Features of hand gestures is used to recognize the letter of sign language using different classification methods such as Hidden Markov Mode (HMM), Artificial Neural Networks (ANN). Sum of Absolute Difference (SAD) algorithm use to extract the image and eliminate the unwanted background noise [8]. The main drawback of vision based sign language recognition system is image acquisition process from video. It has many environmental apprehensions such as the resolution of camera, place of the camera, background condition and lightning sensitivity. Camera is placed such that it captures the maximum focus spot of hand movements, but higher resolution camera required more memory space[21].

Paulo Trigueiros et al. [9] used vision based technique for recognition of Portuguese language. For their implementation, hand gesture was captured in real time. SVM algorithm is used for classification purpose. In this system vowels recognized with accuracy 99.4% and consonants recognized with 99.6% accuracy. M. V. D. Prasad, P. V. V. Kishore, E. Kiran Kumar, D. Anil Kumar [10] presented methods for Indian Sign Language Recognition. wavelet based fusion of two weak edge detection models. One is morphological subtraction model and the other is gradient based canny edge operator. Elliptical Fourier descriptors provide shape models used with optimized number of shape descriptors. Principle components determined find the feature vector to a minimum to accommodate all the frames in the video sequence. Classification of the signs is done by Back Propagation Neural Network Algorithm. The recognition rate stands at 92.34%.

Etsuko Ueda and Yoshio Matsumoto [11] presented a novel technique a hand-pose estimation that can be used for vision-based human interfaces, in this method, the hand regions are extracted from multiple images obtained by a multi viewpoint camera system, and constructing the "voxel Model". Hand pose is estimated. Chan Wah Ng, Surendra Ranganath presented a hand gesture recognition system, they used image furrier descriptor as their prime feature and classified with the help of RBF network. Their system's overall performance was 90.9%. Claudia Nolker and Helge Ritter presented a hand gesture recognition modal based on recognition of finger tips, in their approach they find full identification of all finger joint angles and based on that a 3D modal of hand is prepared and using neural network.El-Bendary et al. developed an Arabic alphabet signs translator with an accuracy of up to 91.3%. Videos are taken of deaf people which convert into text. The features used are rotation, scale and translation invariant. Videos are converted into Frames. For each frame, the distances between three different black pixels are used to construct the feature vector. In the recognition stage, a multilayer Perceptron (MLP) neural network and a minimum distance classifier (MDC) are used [12].

IV. PROPOSED METHOD FOR SIGN LANGUAGE TO SPEECH CONVERSION SYSTEM

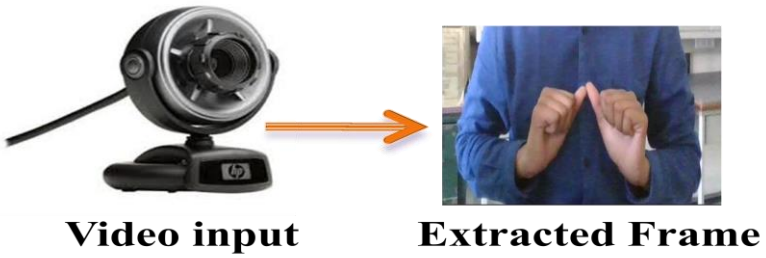


Fig.2. Real time video acquisition of sign hand gestures for ISL.

In this proposed approach, the real time hand gestures for sign language is acquired using laptop webcam with frame size 640×480 and 25 frames per second. The specifications of laptop is HP, i5 processor 8th generation, Ram 8GB with Dedicated graphics.Firstly, the hand movement for sign language for different background condition is captured by using laptop web camera. frame is extracted from real time video for each and every hand movement with frame rate25 frames per second as shown in Fig.2. the accuracy of recognition is depended on how the hand gestures of sign frame is acquired. So that, brightness and orientation of image frame is selected in the specified range to achieve the accurate recognition rate. Brightness of frame is corrected by using image enhancement techniques in preprocessing step. Here, if the orientation of sign hand gestures do not match with specified limit, then there is message in output display of proposed system GUI 'Frameis rejected set the orientation of camera and acquire the video again' as shown in Fig.3. The Extracted frames of sign hand gestures for 26 letters of Indian sign language system from real time video in different background conditions are shown in Fig. 4.

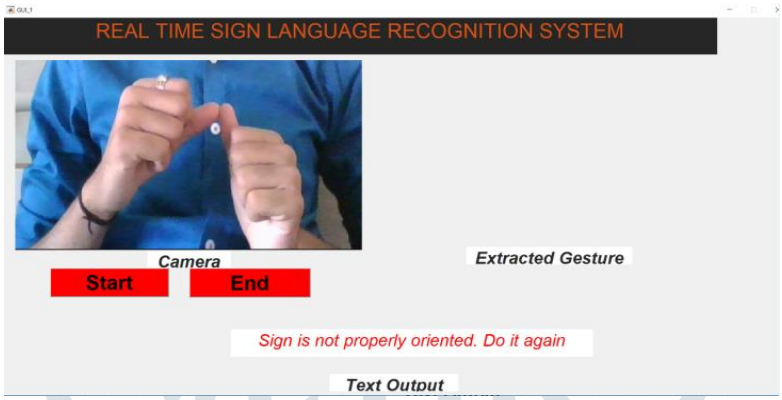


Fig.3. The GUI of proposed sign language system.

Fig.4. Extracted Hand Gestures for Indian sign language system with different background conditions.

3.1 Segmentation of Hand Gesture for Sign.

The hand gesture is segmented from background image by using K-means Clustering algorithm as shown in Fig.5. K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem [21]. The procedure

follows a simple and easy way to classify a given data set through a certain number of clusters. Here, the two clusters are assumed for Hand gesture and background image. The main idea is to define k centres, one for each cluster. These centres should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other.

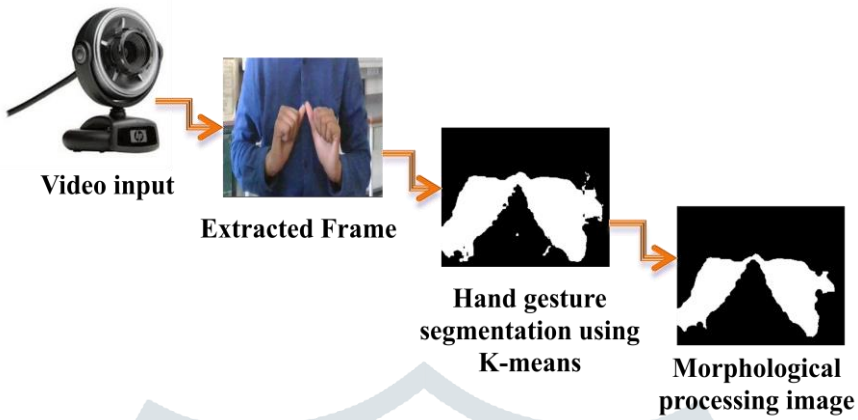


Fig.5.Segmentation of Hand gesture from video input.

The next step is to take each point belonging to a given data set and associate it to the nearest centre. When no point is pending, the first step is completed and an early grouping is done. At this point we need to re-calculate k-new centroids as barycentre of the clusters resulting from the previous step. After we have these k-new centroids, a new binding has to be done between the same data set points and the nearest new centre. A loop has been generated. As a result of this loop we may notice that the k-centres change their location step by step until no more changes are done or in other words centres do not move any more. Finally, this algorithm aims at minimizing an objective function know as squared error function given by:

$$J(v) = \sum_{i=1}^{c}\sum_{j=1}^{c_i}(\|x_i - v_j\|)^2$$

where,
$\|x_i - v_j\|$ is the Euclidean distance between x_i and v_j.
'c_i' is the number of data points in i^{th} cluster.
'c' is the number of cluster centres.

3.3.1 Algorithmic steps for k-means clustering

1. Let x = {x_1,x_2,x_3,……..,x_n} be the set of data points and V = {v_1,v_2,……..,v_c} be the set of centres.
2. Randomly select 'c' cluster centres.
3. Calculate the distance between each data point and cluster centres.
4. Assign the data point to the cluster centre whose distance from the cluster centre is minimum of all the cluster centres..
5. Recalculate the new cluster centre using:

$$v_i = (\frac{1}{c_i})\sum_{j=1}^{c_i} x_i$$

where, 'c_i' represents the number of data points in i^{th} cluster.
6. Recalculate the distance between each data point and new obtained cluster centres.
7. If no data point was reassigned then stop, otherwise repeat from step 3.

By using above algorithm, the hand gesture for 26 letters is segmented. This algorithm segment hand gesture based on skin color, but there is chances to recognize the other unwanted object which as same color intensity like skin. So that erosion of image and some morphological image techniques are applied to remove small unwanted objects in the image. The erosion operator takes two pieces of data as inputs. The first is the image which is to be eroded. The second is a (usually small) set of coordinate points known as a structuring element (also known as a *kernel*). It is this structuring element that determines the precise effect of the erosion on the input image. In this approach line and diamond shaped structuring element are used to remove small objects in the image.

3.2 Recognition of Hand Gesture using Artificial Neural Network

The segmented hand gestures are used to train and test of Artificial Neural Network to recognize the pattern of sign for particular letter. Here, to recognized the pattern, deep learning convolution neural network is used. A CNN uses a system much like a multilayer perceptron that has been designed for reduced processing requirements. In this approach, the layers of a CNN consist

of an 28 input layer, an output layer and a hidden layer that includes multiple convolutional layers, pooling layers, fully connected layers and normalization layers. The removal of limitations and increase in efficiency for image processing results in a system that is far more effective, simpler to trains limited for image processing and natural language processing.

Detection using CNN is rugged to distortions such as change in shape due to camera lens, different lighting conditions, different poses, presence of partial occlusions, horizontal and vertical shifts, etc. However, CNNs are shift invariant since the same weight configuration is used across space. In theory, we also can achieve shift invariantness using fully connected layers. But the outcome of training in this case is multiple units with identical weight patterns at different locations of the input. To learn these weight configurations, a large number of training instances would be required to cover the space of possible variations.

A CNN is trained by using segmented image of hand gesture for 26 letters as shown in Fig.6. After the training of neural network, it is ready to work in real time approach to recognize the sign for 26 letters. The CNN recognizes the sign of letter successfully, recognized letter is converted into text to speech form also it is displayed in text form for user output as shown in Fig. 7.

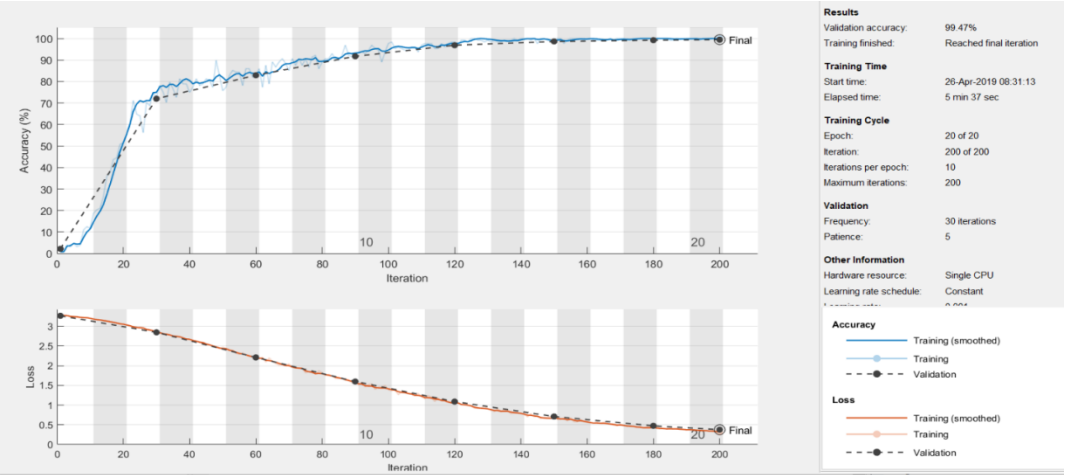


Fig.6. Training of Convolutional Neural Network.

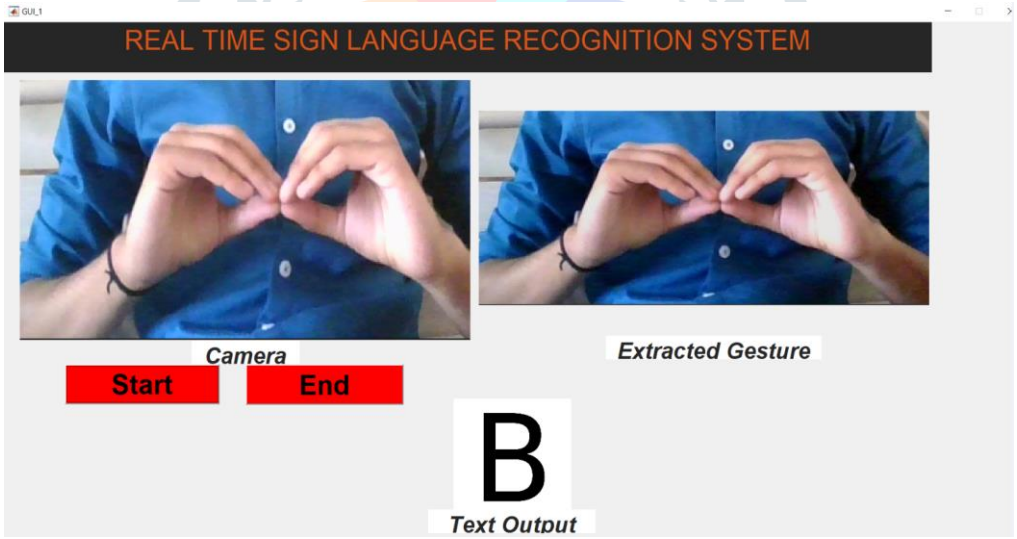


Fig.7. Proposed Real time Indian sign language system.

I. RESULTS AND DISCUSSION

Higher recognition rate of letter is depended on how the frame is acquired and segment the hand gestures. The result of segmented hand gestures from frame of 26 letters of Indian sign language in different background condition is given in Table 1. There is difficulty in segmentation of only sign gestures due to detection of other body parts because of same skin color as shown in Fig. 8. For the sign of letter M,N and R the segmented hand gestures area is almost same. So, the CNN required the number of training data set images are more compared to other letters. The CNN recognizes the all 26 letters of ISL accurately with input layers 28 and number of training data set images are 200. Sometimes also CNN gives false recognition, when the sign gestures of letter M,N, and R is not properly segmented due to its hand gestures are almost same. So, the Convolutional Neural Network gives recognize rate around 97%.

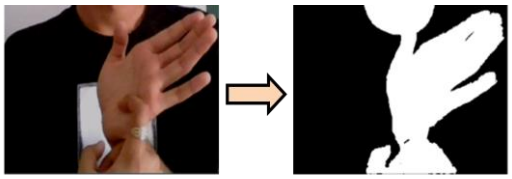
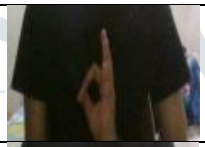
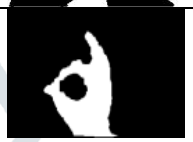
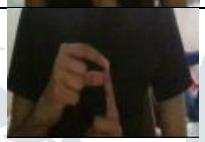
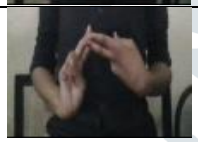
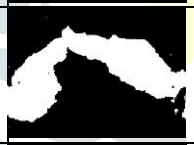
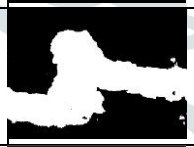
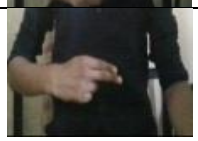
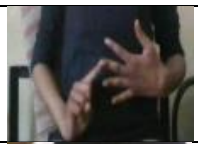
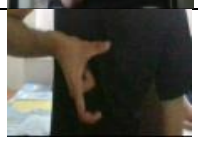
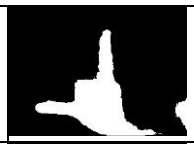
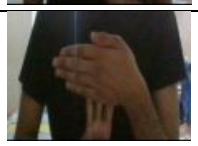
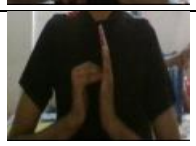


Fig. 8. Segmentation of hand gesture.

Table 4.1: Results of Segmented Hand Gestures for 26 letters of ISL

Letter	Extracted Hand Gesture	Segmented Hand Gesture	Letter	Extracted Hand Gesture	Segmented Hand Gesture
A			N		
B			O		
C			P		
D			Q		
E			R		
F			S		
G			T		
H			U		
I			V		
J			W		
K			X		

L			Y		
M			Z		

V. CONCLUSION

In this paper, we proposed new approach for Indian sign language system using artificial intelligence in real time. When there is the design of system based on image processing, the image acquisition conditions are more important to follow further steps with better accuracy. Here, difficult task is to remove background of hand gestures and accurately segment the sign hand gestures in real time due to other body parts acquired in the image and users do not set proper orientations of web camera. So, that the brightness and orientation of sign hand gestures need to be correct. K-means clustering algorithmgives better accuracy for segmentation for each sign hand gestures of 26 letters. Also, It is depended on what is the speed of users to change the sign for different letters. A Convolutional neural network gives 97% recognition rate for recognition of pattern for Indian sign language system. Sometimes, there is false detection in M, N, and R letter sign, because of they have almost same area of segmented hand gestures. This system gives output in audio and text form so that users can easily communicate.

REFERENCES

[1] Rajaganapathy. , Aravind. , Keerthana. Sivagami. 2015. Conversation of Sign Language to Speech with Human Gestures,2nd International Symposium on Big Data and Cloud Computing (ISBCC'15).

[2] Shreyashi Narayan Sawant1,M. S. Kumbhar2. 2014. Real Time Sign Language Recognition using PCA,IEEE International Conference on Advanced Communication Control and Computing Technologies (lCACCCT),ISBN No. 978-1-4799-3914-5114.

[3] Nicholas Born.2010. Senior Project Sign Language Glove, ELECTRICAL ENGINEERING DEPARTMENT. California Polytechnic State University,1-49.

[4] Kirsten Ellis and Jan Carlo Barca.2012. Exploring Sensor Gloves for Teaching Children Sign Language Advances in Human-Computer Interaction.1-8.

[5] Yun Li, Xiang Chen et al,2012. Sign-Component-Based Framework for Chinese Sign Language Recognition Using Accelerometer and sEMGData",IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 59, NO. 10,2695-2704.

[6] M. Ebrahim AI-Ahdal & Nooritawati Md Tahir. 2012. Review in Sign Language Recognition Systems Symposium on Computer & InfOlmatics(ISCI),pp:52-57, IEEE.

[7] Gunasekaran. K, Manikandan. 2013. R, International Journal of Engineering and Technology (IJET): Sign Language to Speech Translation System Using PIC Microcontroller, ISSN : 0975-4024 Vol 5 No 2.

[8] Wen Gao and GaolinFanga. 2004. A Chinese sign language recognition system based on SOFM/SRN/HMM.Journal of Pattern Recognition".2389-2402.

[9] Paulo Trigueiros, Ferando Ribeiro, Luis Paulo Reis, 2014. Vision Based Portuguese Sign Language Recognition System, New Perspectives in Information System and Technologies,Vollume 1,Advances in Intelligent System and Computing, Springer International Publishing Switzerland,pp-605-608.

[10] M. V. D. Prasad, P. V. V. Kishore, E. Kiran Kumar, D. Anil Kumar,2016. Indian Sign Language Recognition System Using New Fusion Based Edge Operator", Journal Of Theoretical And Applied Information Technology 30th June 2016. Vol.88. No.3, Issn: 1992-8645 Www.Jatit.Org E-Issn: 1817-3195.

[11] Etsuko Ueda, Yoshio Matsumoto, Masakazu Imai, Tsukasa Ogasawara,2003. Hand Pose Estimation for Vision Based Human Interface, IEEE Transactions on Industrial Electronics,Vol.50,No.4,pp.676-684.

[12] Hemina Bhavsar , Dr. Jeegar Trivedi ,Review on Feature Extraction methods of Image based Sign Language Recognition system.

[13] RiniAkmeliawatil, Melanie PO-LeenOoi et al. 2007. Real-Time Malaysian Sign Language Translation using Color Segmentation and Neural Network. Instrumentation and Measurement Technology Conference Warsaw, Poland.IEEE.

[14] Yun Li, Xiang Chen et al,2012. Sign-Component-Based Framework for Chinese Sign Language Recognition Using Accelerometer and sEMGData,IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 59, NO. 10,2695-2704.

[15] Manisha U. Kakde1 , Mahender G. Nakrani2 , Amit M. Rawate3. 2016. A Review Paper on Sign Language Recognition System For Deaf And Dumb People using Image Processing. International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 5 Issue 03.

[16] Suriya M. , Sathyapriya N. ,Srinithi M. ,Yesodha V.2016. Survey on Real Time Sign Language Recognition System: An LDA Approach", International Conference on Explorations and Innovations in Engineering & Technology (ICEIET - 2016), ISSN: 2348 – 8387.

[17] Md. Mohiminul Islam,Sarah Siddiqua and Jawata Afnan,2017. Real Time Hand Gesture Recognition Using Different Algorithms Based on American Sign Language “,978-1-5090-6004-7/ IEEE.

[18] Sérgio Bessa Carneiro, José O. Ferreira, Symone G. Soares Alcalá, Edson D. F. de M. Santos ,Talles M. de A. Barbosa,2016.Static Gestures Recognition for Brazilian Sign Language with Kinect Sensor”978-1-4799-8287-5/2016 IEEE.

[19] Cheok Ming Jin, Zaid Omar, Mohamed Hisham Jaward,2016. A Mobile Application of American Sign Language Translation via ImageProcessing Algorithms, 2016 IEEE Region 10 Symposium (TENSYMP), Bali, Indonesia, 978-1-5090-0931-2.

[20] Hemina Bhavsar , Dr. Jeegar Trivedi 2017. Review on Feature Extraction methods of Image based Sign Language Recognition system, Indian Journal of Computer Science and Engineering (IJCSE),ISSN : 0976-5166, Vol. 8 No. 3.

[21] The k-means algorithm - Notes by Tan, Steinbach, Kumar Ghosh.