International Journal of Research in Advent Technology (IJRAT) Special Issue
E-ISSN: 2321-9637
Available online at www.ijrat.org
*International Conference "*INTELINC 18*", 12th & 13th October 2018*

A Survey on Hand Gesture Recognition Methods for Physically Disabled People

[1]S.G.Praveena, [2]Mrs.C.Jayasri
[1]PG Student, Department of ECE, A.V.C College of Engineering, Mannampandal.
[2]Assistant professor Department of ECE, A.V.C College of Engineering, Mannampandal.
[1]praveena.pravi09@gmail.com,[2] jayasrivijayaraj@gmail.com

Abstract - Communication is the process of exchanging ideas, thoughts, information, and feelings in the form of verbal and non-verbal messages. But it is difficult for deaf and dumb people to communicate with normal people. Thus gestures are the primary method to convey their messages to the normal people and also among themselves. Hence we focus on removing barrier of communication between the disabled people and normal people. This paper reviews on different methodologies for removing this barrier using vision based approaches and non-vision based approaches to convert the sign language into text and audio output.
Keywords- *ASL, Gesture recognition, Wearable device.*

I. INTRODUCTION

Hand gesture is a method of communication between dumb and deaf people. Normal people cannot communicate to dumb people. They need to learn sign language. There are many sign languages are in use. Some of the sign languages are American Sign Language (ASL), Indian Sign language (ISL), Japan Sign Language (JSL), Germany Sign Language (GSL). Here we use Indian Sign Language (ISL). Alphabets used in Indian Sign Language are shown in the figure.

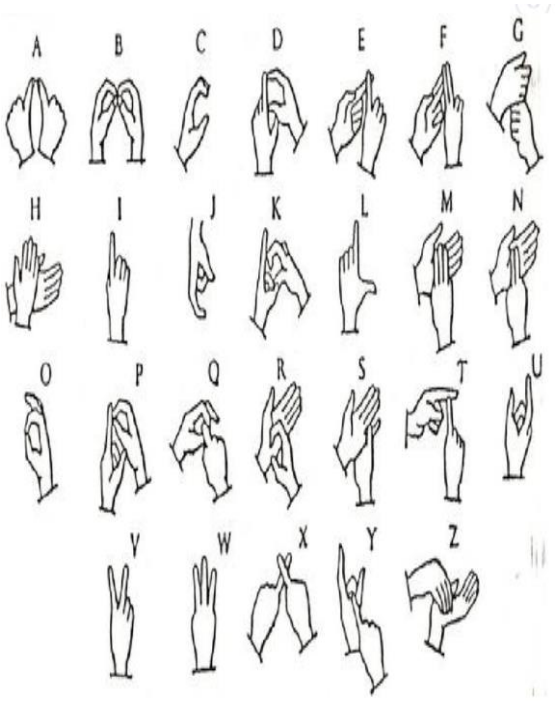

Figure. 1. ISL for alphabets

Hand recognition is done by two approaches known as vision based and non-vision based approaches. In the vision based approach camera plays a major role. It captures the images of the hand gesture and then further processing is done through various techniques. Commonly used approach is the vision based technique. Image acquisition, image preprocessing, feature extraction, classification of the image, and getting final result from the acquired hand gesture are the basic process involved in it. In the non vision based approach sensors are used. Sensors are essential in this type of gesture recognition. Different types of sensors used are flex sensor, tactile sensor, pressure sensor. While using the vision based approach lighting is important for acquiring the image. Proper lighting is required for the vision based approach. There is no need for the proper lighting for sensor based technique.

Sensors are used to sense the different hand movements. Flex sensors are used to measure the degree to which the finger is bent. In the flex sensor, the resistance varies equivalent to the amount of bend applied to the sensor. This resistance is then converted into a voltage value. By comparing the resistance value to the predefined values the output is obtained.

II.RELATEDWORKS

Different methodologies have been used for the sign language interpretation. Camera plays a vital role in the vision based hand gesture recognition. Gestures are captured with the help of the in-built camera. Then the captured image is used for the further process. This process includes removing of unwanted noise, adjusting brightness,

contrast of image, cropping of image, image enhancement, segmentation, feature extraction and sign classification. The image captured in camera was RGB images. Those images were very much sensitive to various light conditions. Hence the image was converted into YCbCr [6]. Color filtering and skin segmentation. In this level RGB format is converted into HSV image. This is helpful to differentiate the skin color and non-skin color.

Images acquisition, image preprocessing are the common steps. Feature extraction differs from one method to another method. Subhankar proposed a system that uses SIFT algorithm [1]. It is used to calculate the source image feature vector. Minimum distance between the feature vector and the source image is found. The alphabet with the highest match in the feature vector is recommended as the output. If there is another sign with the highest number of match difference value is derived. When the difference derived is higher than the threshold voltage then it produces the output "NO MATCH".

In another method [2] of gesture recognition, image is acquired using a camera module that can be directly connected to the controller board. Subsequent steps are preprocessing that include skin color extraction, reducing dimension using blurring, RGB to binary conversion, edge detection. After preprocessing image feature is extracted by different steps like obtaining the contour, calculating the centroid, determining the hull and defect points etc. This gives the corresponding number values by identifying the number of fingers. Finger alphabets can be recognized using matching algorithms from the database created during the training phase. 1- D Hidden Markov model algorithm is used for speech recognition.

Harris algorithm [7] is another algorithm used for the feature extraction. It detects the necessary interest points in the form of Nx2 matrix. Then region of interest is converted to a gray level image on which this algorithm is to be applied for the feature extraction. After the extraction, comes the most important part, i.e. feature matching. The data set already has the features extracted of every standard image stored in the form of a Nx2 matrix in a mat file. The matrix value of the query image is then matched with each of those in the data set of every image. And the minimum distance between the matched features is calculated to get the desired result. The one with the minimum value is the maximum matched image. The result is displayed in the form of text and audio form.

The system described in the paper [9] is implemented using MATLAB. In this approach, firstly, the signs are captured using a webcam. The images captured are then processed further and the features are extracted from the captured images using Principle Component Analysis (PCA). Comparison of the features is done using Euclidean Distance with the training sets. Minimum Euclidean distance helps to recognize the character.

In one of the methods, classification of the ASL hand gestures was done by implementing k-Nearest Neighbor Classifier (kNN). The extracted PCA features are used to classify the ASL alphabets. The generated PCA features are given to kNN classifier. The kNN effectively classifies the ASL hand gestures based on the features.kNN classifier classifies instances based on the similarity to the instances in the training data. The similarity can be determined by computing distance measures to all training data (eg. Euclidean Distance). The k- nearest neighbors are determined, where $k > 1$, neighbors and letting the majority vote decide the outcome of the class labeling.

In another recognition system [20], for skin area detection they used cluster detection algorithm. For boundary filling edge detection algorithm and pixel detection was done by using pixel detection algorithm.

Flex sensors are used to measure the degree to which the fingers are bent. Accelerometer within the gesture recognition system is used as a tilt sensing element, which in turn finds the degree to which the finger is tilted. Tactile sensor is used to sense the physical interaction between the fingers. The outputs from the sensor systems are sent to the Arduino microcontroller unit. In Arduino microcontroller unit, data derived from the sensor output is then compared with the pre-defined values. The corresponding gestures (matched gestures) are sent to the text-to-speech conversion module in the form of text. The output of text-to-speech synthesis system is heard via a speaker [10].

A data glove used by one the systems [12] that consists of five Flex sensors on fingers (Thumb, index, middle, ring, and pinky) and one accelerometer of PIC microcontroller LM386 Speaker Speak jet TTS256 Three outputs (X, Y, and Z positions). Tilting of the palm can be captured by the accelerometer where Flex sensors can measure the bend of the five fingers when making a sign.

When the user performs a gesture/letter and press a button, signals coming from the sensors are amplified via a dedicated amplification circuit to each signal, and then the microcontroller which convert the analog signals to digital values through

its 8-channel ADC. These values are formatted into a simple state matrix: five values for the Flex sensors, one for each axis of the accelerometer. As a result, each letter in the ASL will have a specific digital level for the five fingers and the three axis of the accelerometer.

In the paper [17], they presented a real-time hand gesture recognition system using a combination of image processing modalities. A prototype Graphical User Interface application for ASL sign capture, processing, collection and analysis was presented. The approach consists of a gesture extraction phase followed by a gesture recognition phase. An image gesture database is collected through the application and used as training information to be used in the gesture recognition stage.

Two different translation paradigms: 1) English characters (alphabet) and 2) complete words or phrases. In that method, individual characters, the hand gesture image were processed by combining image segmentation and edge detection to extract morphological information and then processed by the gesture detection stage that recognizes the corresponding alphabet letter. The translation of words and phrases consists of splitting a video sequence into frames and pre-process them in a feature (frame) selection stage. In this feature selection stage, subset of frames that can represent a particular word or phrase are selected. The collection of frames representing a word or a phrase is then processed using the multi-modality technique used for processing individual characters. Finally, the gesture recognition stage is applied to both approaches using a cross-correlation coefficient based scheme to detect the expression.

III.METHODOLOGY

1 .VISION BASED APPROACHES

A.CONVEXITY HULL ALGORITHM

Image based hand gesture recognition is done in three steps. They are image preprocessing, tracking and recognition. The various steps involved in this system [5] are given. The first step in this is acquisition of data from the camera. Then the image preprocessing is done. The preprocessing includes removing unwanted noise, adjusting brightness and contrast of the image, image enhancement and segmentation, color filtering and skin segmentation. The captured image is in RGB format. The RGB is converted into YCbCr and then it is converted into binary image. To get the accurate boundary of the object, the value of HSV is adjusted with the range 0 to255. After this process thresholding is done. Thresholding is the process of separating the image based on the light and dark region.

In contour detection convexity hull algorithm uses for drawing contour around the palm and finger detection. Initial step in convexity algorithm is to segment image in which the hand is located. The maximum and minimum x & y co-ordinate points were computed by joining those points from bounding rectangle. Like hull there are also other points namely convex defects which are present in between valley of two fingers then by taking average of all such defect points surely get a center of palm. So radius of palm consider as a depth of palm only.

Then the ratio of palm radius and distance of hull point from the center point of palm should be more or less depending upon the finger opening and closing position. The points which joint the boundary of hand are called convex hull and the alphabets denotes gap between fingers are called as convex defects.

Hence for recognition of finger position, distance of convex hull (D) from center should be greater than the radius (R) of inner defects circle. Depending on the value of D and R, finger opening and closing can be determined. This algorithm is very convenient for finger point detection and number recognition. This system is suitable only for number recognition.

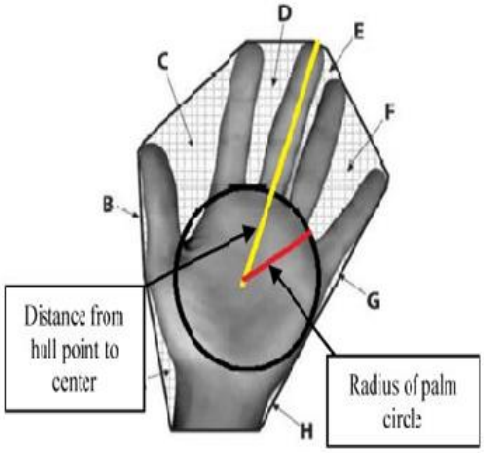

Fig.2 .Convex hulls and defects

B.SIFT ALGORITHM

The system proposed [1] has been tested and it is faster than the earlier system. The system works on three main features that is image Pre-processing, features extraction using scale-invariant feature transform (SFIT) algorithm and lastly stratification of the image. In image preprocessing we have to construct a database with different gestures as used in ASL. We also have to decrease

the distortion and resizing of the image has to be done because usually MATLAB uses default function the user needs to resize it according to individual need. In order to do this we will use bi-cubic method to resize the input image. Bi-cubic method maps the targeted pixels to the sixteen nearest sources. We will also use a low pass filter to reduce the distortion in the image.

The next step is segmentation. Segmentation of gesture involves extraction of the gesture from the image background.

The segmented image usually contains distortion or noise but the distortion is reduced in our system by the use of low pass filter in the image pre-processing stage. Firstly all possible scales of the image are searched at all locations of the image using the scale space [9] and identification of PIP (potential interest points) is done using the Gaussian function and difference of Gaussian is given. Secondly the key point localization where the most stable possible set is selected and escalated for further input image localization. Thirdly is the orientation that is assigning to the key points and fourthly the key point caption. $L(X,Y,\sigma)$ is expounded as the scale space of the image, $G(X,Y, \sigma)$ is derived from developed from the adding of irregular scale Gaussian along with the image input, $A(X,Y)$.

The maxima and minima of the image are then located by checking each pixel. On the basis of the image feature the orientation is given to the interest points which are stable. In the key point caption a 128 bytes forms the feature vector. We implement SIFT algorithm to calculate the source image feature vector. Minimum Euclidean distance between each feature vector of the source image is found.

Then all the feature vector of the source image is compared with the ASL sign in the database. The alphabet with the highest number of match in feature vector is recommended but there is another sign with second highest number of match. So the conflicting sign with the recommended sign is compared and the difference between them is derived. If the difference is a larger value compared to threshold than the recognition of the image is successful in equating to the highest recommended image but if the difference is of less value than the threshold then recognition is not done or "NO MATCH" is given as output.

2. NON-VISION BAESD APPROACHES

A.SVM CLASSIFIER

A smart sign language interpretation system using a wearable hand device is proposed. This wearable system utilizes five flex-sensors, two pressure sensors, and a three-axis inertial motion sensor to distinguish the characters in the American Sign Language alphabet. The entire system mainly consists of three modules: a wearable device with a sensor module and a processing module, and a display unit mobile application module.

The finger gestures are observed through the flexion of the flex sensors, whereas the hand gestures are examined based on the hand motion through the orientation derived from an inertial motion sensor.

The flexion degree is split into three regions in each vector. The first region is denoted as "no bend" or "slight bend," which is associated with a normalized flexion value within the range of [0.0, 0.3). The second region is considered as a "partial bend" with the associated normalized flexion value within the range of [0.3, 0.7), and the last region is a "complete bend" with associated normalized flexion value within the range of [0.7, 1.0]. These regions are abbreviated in order as OR (open region), PR (partially open or closed region), and CR (closed region). Table 1 shows a mapping of these regions for all 26 alphabet letters in ASL, as well as a "neutral" gesture.

A 9-degree-of-freeom (9-DOF) IMU is integrated with an MEMS accelerometer, magnetometer, and gyroscope under a single die, and processed using a high-speed ARM Cortex-M0 processor. The device abstracts the sensor fusion and derives the orientation data in quaternion, Euler angles, or in a vector format. The flex sensor and IMU data are collected using an Arduino Pro Mini 328.

Then the features are extracted from the sensor data and serve as inputs to the built-in SVM classifier to determine the sign language alphabet letters. In this system [15], there are a total of 28 gesture patterns, which refer to the 26 alphabet letters of ASL, a "neutral" state (indicating no gestures to be observed), and an invalid sign.

The pressure sensors are used to differentiate the alphabets which are close to each other. It produces the output 1 if it is pressed and it produces the output 0 if not pressed. In this, the signs are classified into 28 classes using a support vector machine (SVM). Now, the detected sign is translated into text and transmitted to a mobile device using a Bluetooth 4.0 module. Text-to-speech service is also implemented in the application.

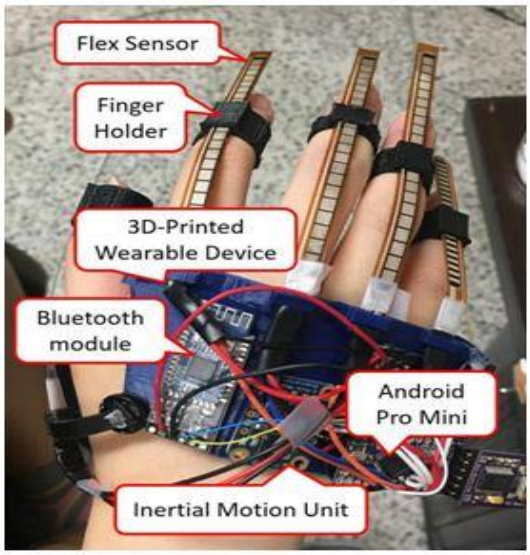

Fig.3.3D printed wearable device that holds the hardware components, which include an Android Pro Mini microcontroller, a flex sensor, a motion sensor, and a Bluetooth low energy (BLE) module.

B.HMM ALGORITHM

The flex sensors are mounted on the glove and they are fitted along the length of each of the fingers. Depending upon the bend of hand movement different signals corresponding to x-axis, y-axis and z-axis are generated. Flex sensors outputs the data stream depending on the degree and amount of bend produced.

The output data stream from the flex sensor, tactile sensor and the accelerometer are fed to the Arduino microcontroller, where it is processed and then converted to its corresponding digital values. The microcontroller unit will compare these readings with the pre-defined threshold values and the corresponding gestures are recognized and the corresponding text is displayed. The text output obtained from the sensor based system is sent to the text-to-speech synthesis module.

Hidden Markov models (HMMs) is the most popular speech synthesis technique. The output produced is In HMM-based speech synthesis, the speech parameters such as the spectrum of speech signal, its fundamental frequency (F0), and the duration information for the phonemes are statistically modeled and the speech is synthesized using HMMs based on the maximum likelihood criteria. The model is parametric because it describes the speech using parameters and it is statistical because the parameters are described using means and variances.

To eliminate the spectral discontinuity in the spectrum of synthesized speech along with static coefficients, delta and acceleration coefficients are also computed. Then the system will convert the gesture to the corresponding text and then the speech is synthesized for the corresponding text by using the text-to- speech synthesizer.

IV.RESULT

When compared to the vision based algorithm, non-vision based algorithm produces accurate result. But the non-vision based approach is suitable only for alphabets.The result produced by the above mentioned methods are shown below

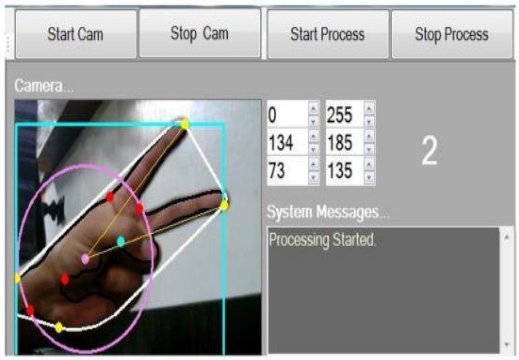

Fig.3.Recognition result

The fig.3 shows the result of vision based approach using convexity hull algorithm. As this system is vision based proper lighting is required to capture the image. The hand gesture recognition system proposed in the paper [1] uses SFIT algorithm for feature vector implementation and Euclidean distance for vector comparison. ASL is recognized through sign images which will help in communication with people who do not possess the knowledge of it.

The SFIT algorithm has been used for reducing the delay time and increasing the accuracy in results. This system also exhibits the distortion reduction characteristic in the images so that the accuracy of the recognition can be maintained.

COMPARISON OF CLASSIFICATION RESULTS FOR SIGN RECOGNITION FOR ALPHABET 'R', 'U' AND 'V'

Alphabet \ Subject	R	U	V
1	65.1 / 97.2	66.2 / 97.5	65.8 / 98.5
2	51.2 / 96.4	52.1 / 97.1	50.5 / 98.2
3	53.4 / 97.4	52.8 / 98.6	53.5 / 97.2
4	56.1 / 95.5	55.2 / 96.4	55.3 / 97.5
5	61.5 / 95.8	60.8 / 97.5	61.1 / 96.7
6	62.8 / 97.9	61.8 / 97.8	60.5 / 96.9
7	60.8 / 96.4	61.1 / 98.4	61.0 / 97.5
8	69.5 / 98.7	68.4 / 96.7	67.5 / 98.5
9	51.6 / 97.5	50.8 / 97.2	51.1 / 96.9
10	52.7 / 98.6	51.8 / 97.8	50.8 / 97.6
11	53.8 / 99.5	52.8 / 99.4	53.1 / 98.2
12	54.8 / 96.4	53.2 / 95.8	54.2 / 96.1

Fig.4.Comparison of accuracy percentage without/with using pressure sensor

In sensor based technique results depend on the sensor and the component used in that particular system. The accuracy percentage of using flex sensor alone and the flex sensor with the pressure sensor used in the two systems are shown in the above figure. This wearable system with the fusion of pressure sensors on the middle finger increased the recognition accuracy rate dramatically to 98.2%.

REFERENCES	METHODOLOGY	LIMITATIONS
[3]	Gestures are taken from the camera and then they are classified with the help of KNN classifier The recognized gesture is compared with the database and the corresponding output is produced.	Only 20 alphabets are recognized out of 24
[7]	This system also uses vision based technology. It uses harris algorithm for hand gesture recognition. This system works properly even when the hand is in contact with the body.	Both the hands cannot be used. Very light and very dark background may affect the result.
[8]	Hand gesture recognition is fully based on android application. It uses java and android studio. By using this application accuracy is more.	It detects the gesture only from a fixed position. Application size is large because it stores images of gestures used for text to sign feature.
[2]	ARM Cortex processor is used for hand gesture recognition. Speech recognition is carried out by using HMM algorithm. Power consumption is low for this system	In speech recognition part, background noise may affect the accuracy of the result.
[10]	This system is implemented using MATLAB and the feature extraction is done using PCA	Accuracy depends on the white colored background.

Table 1.Comparison of various methodologies

CONCLUSION

In this paper, the concept of different methodologies for conversion of gesture into text and audio output are discussed and compared. In the vision based technique lighting is essential. Without proper lighting we could not able to find accurate result. While using sensor based system there is no need for proper lighting. But the sensor based technique is suitable only for alphabets. By using vision based approach we can expand the system to word recognition. This system is also economical.

REFERENCES

[1] Subhankarchattoraj, Karan vishwakarma, Tanmay Paul, "Assistive system for physically disabled people using gesture recognition" presented at 2017 IEEE International conference on signal and image processing.

[2] Geethu G Nath, Anu S, "Embedded sign language interpreter system for deaf and dumb people" presented at 2017 International Conference on Innovations in information embedded and communication systems (ICIIECS).

[3] Sruthi upendran,Thamizharasi A, "American sign language interpreter system for deaf and dumb individuals" presented at IEEE International conference on control, instrumentation communication and computational technologies(ICCICCT).

[4] Mrs.NeelaHarish, Dr.S.Poonguzhali, "Design and development of hand gesture recognition system for speech impaired people", 2015 IEEE International conference on industrial instrumentation and control (ICIC).

[5] AbinandanDas,Lavish Yadav, Mayank singhal,Ramansachan, Hemanggoyal, Keshav taparia Raghav Gulati,Ankitsingh, Gaurav Trivedi, "smart glove for sign language communications",2016 International conference on Accessibility to digital World(ICADW)

[6] Ashish S. Nikam , Aarti G.Ambekar, "Sign language recognition using image based hand gesture recognition techniques",2016 online International conference on Green Engineering Technologies (IC-GET).

[7] Anchalsood,Anju Mishra "A communication system for deaf and dumb", 2016 5th International conference on Reliability ,Infocom technologies and optimization (ICRITO).

[8] Zulfiar A. Menon , M. Uzair ahmed ,S. Talha hussain, Zahid, "Real time translator for sign languages", 2017 IEEE International conference on Frontiers of information technology.

[9] NiteshS.Soni,prof.Dr.M.S.Nagmode,Mr.R.D.Komati, "Online hand gesture recognition & classification for deaf & dumb"

[10] Aarti M, Vijayalakshmi P, "Sign language to speech conversion", 2016 IEEE 5th International conference on recent trends in information technology.

[11] Piotr Muzyka,Marekfrydrysiak, Elizabetaroszkowsa "Real time detection of hand gestures",2016 International conference on Methods and Models in Robotics.

[12] Mr.kunal A. Wankhade, prof. GuauriN.zade, "Sign language recognition for deaf and dumb people using ANFIS", IJSERT,volume 3, issue 5, may 2014.

[13] V.Padmanabhan, M.Sornalatha, "Hand gesture recognition and voice conversion for dumb people", IJSER, volume 5, issue 5, May 2014.

[14] Manisha U. Kakde, Amit M. Rawate, "PCA" , IJESC volume 6, July 2016.

[15] B.G.Lee, S.M.Lee, "Smart wearable hand device for sign language interpretation system with sensor fusion",IEEESensor journal 2018,volume 18,issue 3.

[16] Anshal Joshi, Heidy Sierra and Arzuaga, "American sign language translation using edge detection and cross correlation", IEEE 2017.

[17] Shital Pradeep Dawane ,Hajiali Ali Sayyed, "Hand gesture recognition for deaf and dumb people using GSM module",IJSR,Vol 6 issue 5, May 2017.

[18] ChanniahChandana K, Nikitha K, Nikitha P, Bhavani N K, Sudeep J, "Hand gesture recognition for deaf ,dumb and blind people", IJIRCCE, vol 5, issue 5, May 2017.

[19] Aswathy M, Heera Narayanan, Surya Rajan, Uthara P M, Jeena Jacob, "Hand gesture recognition for deaf and dumb people using feature extraction", IJAREEIE, vol.6, issue 3, march 2017.

[20] Sarika Turkar, Mrinalini Peter, PayalKapakar, Reshma Dehankar, Hemlata P Sahu, "Text generation using hand gesture recognition", IJIFR, Vol.4, issue, March 2017.