2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS)

Indian Sign Language Translator Using Gesture Recognition Algorithm

Purva C. Badhe
Assistant Professor, Biomedical Engineering Department
D.J.S.C.O.E. Mumbai University
Mumbai, Maharashtra, India
purva.pcb@gmail.com

Vaishali Kulkarni
Professor and Head of EXTC Department
M.P.S.T.M.E. NMIMS University
Mumbai, Maharashtra, India
Vaishali.kulkarni@nmims.edu

Abstract— **Sign Language is a natural language which deaf community uses for communication. Sign Language (SL) is a subset of gestures or signs made with fingers, hands, arms, eyes, and head, face etc. Each gesture in SL has a meaning assigned to it. Understanding SL is nothing but understanding the meaning of these gestures. There exists a problem in communication when a person who completely relies on this gestural SL for communication tries to converse with a person who does not understand the SL. Every country has its own developed SL. In India, this language is called as "Indian Sign Language (ISL). This paper aims to develop an algorithm that will translate the ISL into English. This paper has implemented a system named as "Indian Sign Language (ISL) Translator using Gesture recognition algorithm". The system translates gestures made in ISL into English. The gestures that have been translated include numbers, alphabets and few phrases. The algorithm first performs data acquisition, then the pre-processing of gestures is performed to track hand movement using a combinational algorithm, and recognition is done using template matching. The database used for implementation has been self-created and includes total 130,000 videos; out of which 72,000 videos were used to create the system database and remaining 58,000 videos have been tested for checking the performance of the system. The accuracy of this system is as high as 97.5%**

Keywords— Gesture Recognition; motion detection; hand tracking; segmentation

I. INTRODUCTION

A. Introduction to Gesture Recognition

Gesture as defined by dictionary means 'Motion of limbs or body made to express thought or to emphasize speech'. Gestures are meaningful manoeuvers that involve physical motion of different body parts such as hands, fingers, arms, head, neck, eyes etc. Gestures convey meaningful information. In order to extract information from the performed gestures a gesture recognition system can be implemented. Gesture recognition is the process by which gestures made by the user are recognized by the receiver [1]. The gestures are primarily divided into two classes: Static gestures and Dynamic Gestures. Static gestures include only poses and configurations whereas dynamic gestures include strokes, pre-strokes, postures and phases. The dynamic gestures often include movement of body parts. It may also include emotions depending on the meaning that gesture conveys. Apart from movement phenomena, the second distinguishing factor between static and dynamic gestures is the inclusion of emotions. The dynamic gestures incorporate emotions. For example, to indicate that "The Mountain was big" one can make use of only static signs while if it is to be expressed as "The Mountain was *this* big", one will need to use movements of arms that can be categorized under dynamic gestures [2]. Depending on the context, the gesture may be widely classified as:

- Arm gestures
- Facial / Head gestures
- Body gestures

Typically the meaning of gestures is dependent on [1]:

- Spatial information: where it occurs
- Pathic information: the path it takes
- Symbolic information: signs it makes
- Affective information: emotional quotient

B. Gesture Recognition System

First step of gesture recognition system is to acquire gestural data. There can be various ways to capture this data. The methods are basically classified into two: sensor based method and vision based method. In sensor based method, data glove or motion sensors are included from which the gestural data can be extracted. Vision based method includes image processing. The glove trackers are better in a way that they capture even minute details of the gesture enhancing the system performance. More the number of sensors ensure the better data capturing. However this method requires wearing a hand glove with embedded sensors which make it a cumbersome device to carry. This reduces user comfort. Vision based methods have a benefit that they can include texture and colour parameters to recognition which is difficult to achieve in sensor technology. With image or video processing techniques, 2D and 3D study of gestures can be incorporated depending on the requirements. These systems may encounter occlusion problem, where certain part of the gesture movement gets occluded with other body part. Thus one needs to be very specific of the aim before developing the gesture recognition system for a particular application. Each technology varies in several dimensions namely, accuracy, resolution, latency, range of motion, user comfort and cost [1]. In general a gesture recognition system is as shown in Fig. 1.

978-1-4673-7437-8/15/$31.00 ©2015 IEEE

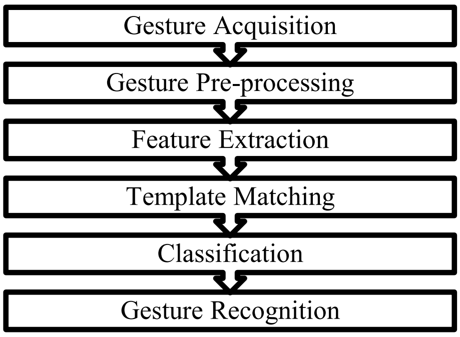


Fig. 1: Generalized block diagram of Gesture recognition system

It includes:

- Gesture Acquisition Block: It is a sensing device which is responsible for capturing the gesture. In case of vision based approaches it is a camera and in case of sensor based approaches it is a data glove, or motion tracker or marker.
- Gesture Pre-processing Block: This block is mainly responsible to make gestural information useful information in feature extraction point of view. This block will enhance the useful data and get rid of unwanted data.
- Feature Extraction Block: It is the block that collects all the feature components of the gesture and stores them in a code vector.
- Template Matching: In this block, the code vector is compared with the existing codebook vectors that are reference vectors in database.
- Classification: Based on the output of template matching, a classification of gestures will be done. This block will classify the gesture as per the nearest match found in template matching.
- Gesture Recognition Block: The terminal block will recognize the gesture completely and produce an appropriate output.

C. Applications of Gesture Recognition System

Gesture recognition has a broad range of applications such as the following [2]:

- Human Computer Interface: Gesture recognition can provide a smart human computer interface by means of Virtual Keypad, Head or eye blink controlled mouse, Gesture controlled projectors and so on.
- Video gaming: by incorporating gestures in video gaming as user interface, as head movement, finger pointing or complete body movement, it will enhance the gaming experience.
- Gestures can provide input to computer when a severely disabled person or very young children want to access computer.
- Gesture recognition will help significantly in case of polygraph; lie detection.
- Augmented Reality: Gestures are useful in navigation or manipulation in case of virtual reality.
- Advanced smartphones can deploy several functions of this technology.
- This technology will be helpful in healthcare industry to diagnose and monitor stress level.
- To interpret Sign Language.
- Security and biometric applications.
- Driver alertness system and so on

II. Indian Sign Language

A Sign Language (SL) is the natural way of communication of deaf community. More than 2.5% population of the world is deaf [3]. According to the studies of The All India Federation of the Deaf, in India around 4 million are deaf and more than 10 million people are hard of hearing [4]. Every country has its own Sign Language developed with a number of grammatical differences. The well-known sign languages are namely [5]:

- American Sign Language (ASL)
- British Sign Language (BSL)
- Spanish Sign Language
- Israeli Sign Language
- Indian Sign Language (ISL)
- Pakistani Sign language
- South Korean Sign Language
- Taiwan Sign Language
- Turkish Sign language
- Arabic Sign Language and so on

More than 1 million of adults and 0.5 million children in India make use of Indian Sign language [3]. The linguistic studies of ISL began from 1978 onwards. ISL is a completely natural language with its own grammar, syntax, phonetics and morphology. ISL provides linguistic information which makes use of hands, arms, face, eye and head and face posture. To convey the meaning, a signer often makes use of a 3D space around her, called Signing Space. ISL contains both, manual and non-manual components. Parameters like shape, orientation, position, and movement of hands characterize a manual component whereas non-manual components are characterized by facial expressions, eye gaze, and head/body posture [6]. In the later decade of 2000 the research began on ISL recognition system. ISL recognition system is a way by which Indian Sign language can be decoded and interpreted in the local spoken language. In order to develop such systems, large database is required. Western countries like the United States of America, United Kingdom have prepared their Sign Language database which is available over internet. However there is no authentic database of ISL available. Hence those who want to work on development of ISL recognition system need to create their own database of gestural signs. The linguistic studies of ISL began from 1978 onwards. ISL is a completely natural language with its own grammar, syntax, phonetics and morphology [4]. Gesture Recognition system developed in [7] examined the input gestures for match with a known gesture in the gesture database. Gesture Database Contained the necessary information required for pattern matching as well as a gesture-to-text dictionary and Speech Synthesis Module Converted word or letters obtained after gesture analysis into corresponding sound. Other different ISL recognition systems are based on binary image dynamic loading technique [8] and the YCrCb skin color modeling approach [9], PCA (Principle Component Analysis) based [10] and a SIFT (Scale Invariant Feature Transform) based system [11]. Table 1 illustrates the comparison of three methods used for ISL Recognition

Table 1: Comparison of different ISL Recognition approaches [3]

Method	Accuracy (%)	Suitable applications	Shortfalls
Vision Based Method	98.125	Single user Environment	Suitable for two handed gestures
YCbCr skin colour Approach	92.7	Real time application with good segmentation	Less accurate during Dynamic gestures
PCA Approach	94	Occluded and overlapped gestures can be recognized. Better for two handed approach.	Accuracy need to be enhanced
SIFT Approach	93	Suitable for two handed gestures	Complex algorithm

III. ISL TRANSLATOR USING GESTURE RECOGNITION ALGORITHM

This section focuses on the flowchart of the proposed system. The system is implemented to perform a task of interpreting the gestures and to decode them. The decoded or in other words, the translated gestures are expressed in English. Flowchart described below in Fig.2 illustrates the complete outline of the workflow of proposed system. The complete system is divided into two parts viz. Training and testing.

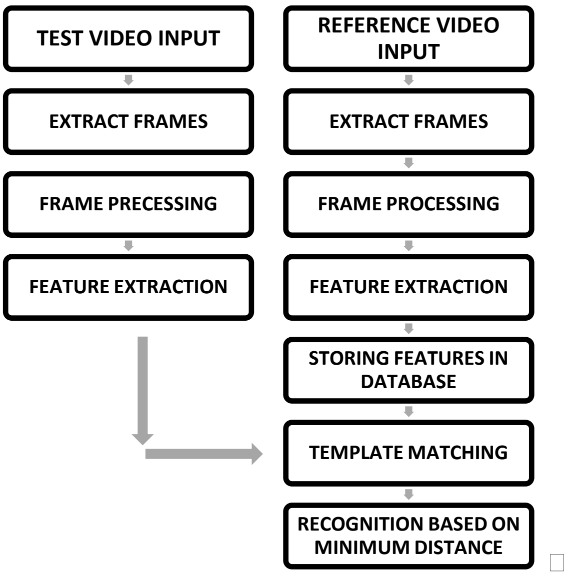


Fig. 2: ISL Translator using Gesture Recognition System outline [12]

A. Training Phase

In training sequence, the workflow of proposed system is as follows:

- The reference video is extracted into frames and the individual frame is pre-processed. In this pre-processing, several filters are applied so as to enhance the useful content of the frame information and to reduce the unwanted information as much as possible.
- All the features of processed frames are then extracted using Fourier descriptor method of feature extraction and stored in database associated with that gesture.
- The same procedure is followed for all the gestures to be included in the system. And the complete reference database is prepared.

B. Testing Phase

The workflow of testing sequence can be outlined as below:

- The input gesture is in terms of video which is extracted for frames.
- The frames are pre-processed in order to get the 'Processed frame' as similar to that in training sequence.
- After obtaining the 'Processed Frame' in testing, the frames are temporarily stored in temporary file and then are matched with the previously maintained database.
- The difference between them is measured depending on minimum distance method. On finding the nearest match, the gesture is recognized as that particular sign and corresponding output is flashed on the screen.

IV. IMPLEMENTATION

A. Data Acquisition

The database of the gestures is self-created. Hence system begins with data collection. Data is acquired from signers who are deaf and mute by birth. As SL is their natural language of communication, data obtained from them is considered to be the most authenticated. The video recording is done at "Ali Yavar Jung National Institute for Hearing Handicapped, Bandra, Mumbai" [13]. The signs recorded include the numbers, alphabets and some phrases. The complete database includes total 130,000 videos out of which 58,000 videos are tested for checking the performance of the system.

B. Pre-Processing and Hand Tracking

The developed system makes use of a combination of three methods [12] for feature extraction. That makes a system to provide a real time hand tracking extraction to trace the moving hand against stationary background and extract a region of gesture [14]. However before performing feature extraction, first the gestures must be handled in such a way that they can produce the most useful information neglecting the redundant, meaningless noise and superficial information that does not contribute in a vital manner to the feature extraction. For this purpose, before proceeding for feature extraction, first the raw data of gestures is trained to give out only useful information. That is achieved using the system shown in following flowchart shown in Fig. 3 which gives us the output termed as 'processed frames'.

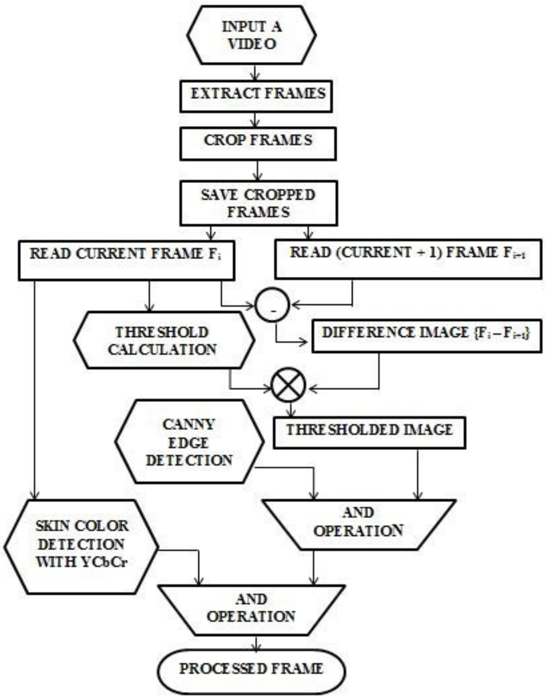
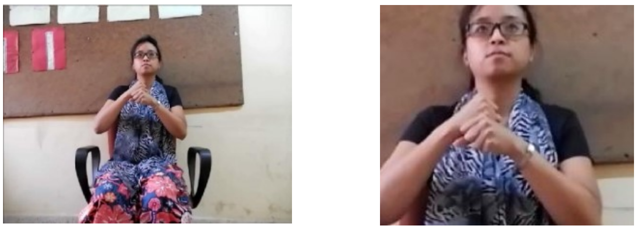


Fig. 3: Flowchart for obtaining Processed frames [12]

Fig. 4 shows the output of image cropping. Cropping is performed manually by giving the crop coordinates in terms of [Horizontal right position, Horizontal left position, Vertical up position, Vertical down position].



(a) (b)

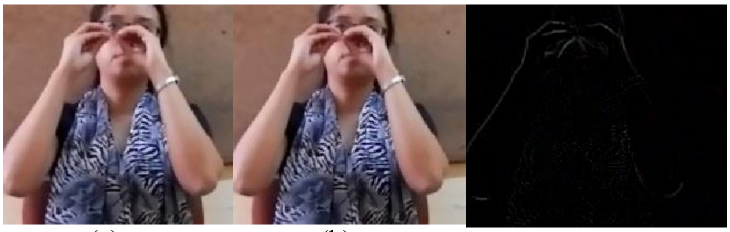
Fig 4: Output of image cropping: a) Original image of gesture indicating alphabet G, b)Cropped output

After cropping the frames, two subsequent frames are subtracted from each other for obtaining the difference image as shown in Fig. 5. In order to achieve segmentation of threshold image, Eq. 1 is implemented on the difference image.

$$Ti = T\{Fi - F(i+1) \qquad (1)$$

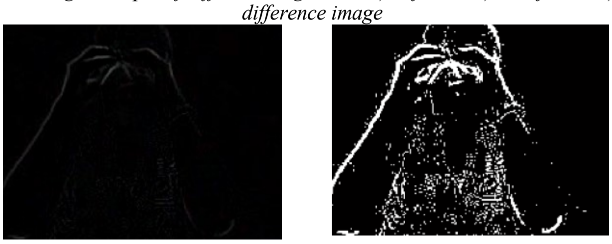
Fig. 6 shows the output after segmentation. Also simultaneously for motion tracking, skin colour detection algorithm is employed using YCbCr model.

Fig. 7 illustrates the output of skin colour detection. The third factor of combinational algorithm is boundary detection. This is achieved using edge detection using canny edge detector.



(a) (b) (c)

Fig 5: Output of difference algorithm. a) i^{th} frame; b)$i+1^{th}$ frame; c) difference image



(a) (b)

Fig 6: Output of segmentation: a) Difference image b) segmented image

(a) (b) (c)

Fig.7: Output of skin colour detection; a) Original cropped image, b) grey scale converted image, c) skin colour detection output using YCbCr model

Fig. 8 shows the output of edge detection. Once the pre-processing of gesture is done, hand motion is traced. This step is important because gesture has movement of hand as an integral part of it. If one is successful in tracking the hand movement, the maximum part gesture is automatically traced. In order to track the hand movement a combination algorithm [12] is used which takes into consideration hand motion, skin colour and the boundary of hand of the signer.

(a) (b)

Fig 8: Output of edge detection a) Original cropped image, b) edge detected image using canny edge detector

Fig. 9 illustrates the combined output of hand tracking.

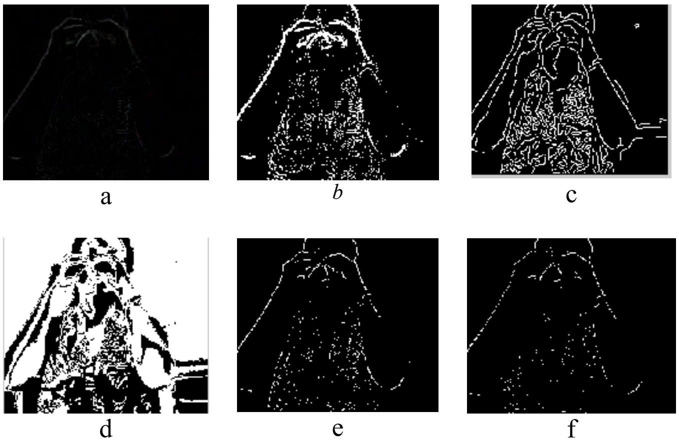


Fig. 9: Output of Pre-processing and Hand Tracking: (a) Difference Image, (b) Output of segmentation, (c) Output of Canny edge detector, (d) Output of Skin colour detection, (e) Combined output of Segmentation and Edge detection, (f) Combined output of Segmentation, Edge detection and Skin colour detection.

C. Feature Extraction

Feature extraction is an important function for pattern recognition. It achieves dimensionality reduction. In case of gestures, the input data i.e. in terms of frames, is too large to be processed and can have redundancies. That is why; input data has to be transformed into its reduced form. The transformation has to follow a set of features. Thus the process of feature extraction represents the large set of data accurately in a simplified manner. There can be several features that can be selected for this purpose. Some approaches make use of Principle Curvature Based Region (PCBR) and wavelet Packet Decomposition (WPD) [9], Principle Component Analysis (PCA) [10], or Scale Invariant Feature Transform (SIFT) [11] method. Fourier descriptors [14] is another successful approach used for feature extraction purpose. This paper uses Fourier Descriptor (FD) method of feature extraction.

1) Feature Extraction using Fourier Descriptors

To extract the external boundary points of a hand shape the contour following algorithm [15] is used. To represent the boundary points, first the Fourier series were calculated using Fast Fourier Transform (FFT) algorithm. For feature extraction purpose, 28 Fourier descriptors per each frame of video gesture are considered. Before selecting 28 descriptors, the system was implemented using 30 descriptors as well as 25 descriptors. The best results were obtained using 28 x 28 descriptors.

2) Codebook Creation

Aim of the training system is to generate a codebook of database. After implementing a feature extraction using FDs the extracted data that was obtained was too large. In order to store the reference codebook, compression was required. This compression has been achieved using vector quantization. Vector quantization (VQ) is a non-uniform and many-to-one mapping [15], responsible for lossy compression. It is based on equivalence relation. Many vectors are mapped into a single region as they are equivalent in some sense [15]. The vector quantization scheme can be divided into three parts: the codebook generation process, the encoding procedure and the decoding procedure [16]. The most important factor of vector quantization technique is to design a good codebook. The type of vector quantization that is employed is popularly known as Linde- Buzo- Gray (LBG) type of vector quantization.

D. Template Matching

Once the Feature extraction is done, the next step is to match the extracted features of a testing sequence with the previously saved training reference codebook vectors. This system makes use of a simple Euclidean Distance method. The Euclidean distance between two points $x=[x1, x2]$ and $y= [y1, y2]$ is given by Eq. 2.

$$d_{xy} = \sqrt{(x1 - y1)^2 + (x2 - y2)^2} \qquad (2)$$

E. Gesture Recognition

By employing formula mentioned in Eq. 9, the distance between testing code vector and each vector from reference codebook is calculated. After calculating all the distances, the code vector which gives the minimum distance with the testing code vector is considered as a match and the corresponding signer's gesture output is displayed on the screen.

V. RESULTS

The system was tested on 10 different signers. The accuracy was checked as per correctness of every gesture made i.e. Alphabets, Numbers and Phrases.

Fig. 10 shows the accuracy of ISL recognition system performance in case of each and every alphabetic gesture. The maximum accuracy is 100% for most of the alphabets with only few exceptions. This implies that the system works efficiently for most of the alphabetic character recognitions. It is observed that accuracy is less in case of alphabets: C, M, O, R and X. This occurs because gesture for sign C and O are less different than each other. Similarly gesture for alphabet M, N and R is very close to each other.

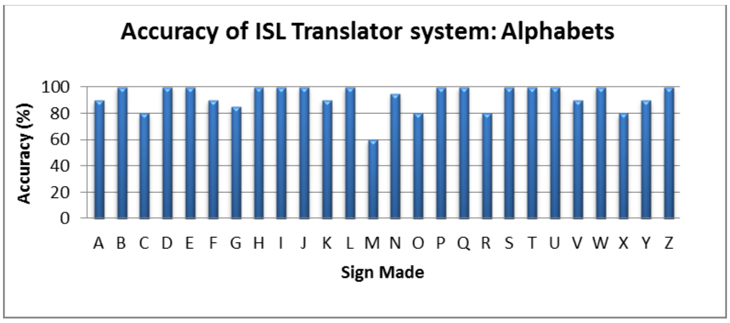


Fig. 10: Accuracy based on Individual Alphabet gestures

As seen in Fig. 11 in most of the cases accuracy is 100% for 4 signs, 95% for 3 signs and 90% for rest. The accuracy does not fall below 90% in any case. In case of the gestures with 90% accuracy, the moving body part is either occluded or tilted.

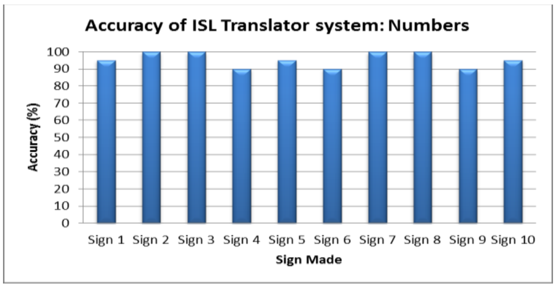


Fig. 11: Accuracy based on Individual Numeric gesture

Fig.12 describes the accuracy of phrase recognition. It can be seen from the graph that out of 10 different signs, 8 signs provide 100% accurate results and out of rest two, one provides 90% and the other provides 85% correct result. The phrases that are less accurate are namely 'Thank you' and 'Fire'. This happens because the gestural movement for 'Thank you' is similar to 'Please' and 'Fire' is similar to 'Danger'

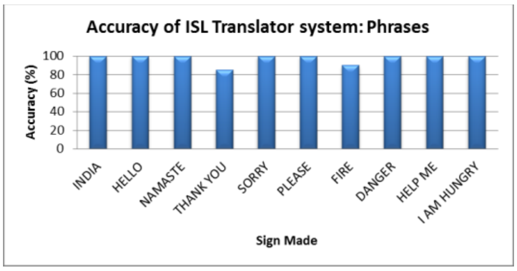


Fig. 12: Accuracy based on Phrase gestures

VI. CONCLUSION

In this paper, a system has been developed that translates the gestures made in Indian Sign language (ISL) into English. The system aims to help the society in terms of better communication aspects. ISL translator system helps in such a way that, it inputs the ISL sign gesture, decodes it, interprets it and outputs the meaning of the gesture in English. The database for developing this system is created on own with the recorded videos of deaf and mute signers. This makes the gestures included to be authentic. The database is created using 78,000 different videos of total 130,000 recorded videos. In order to create database, the input videos are converted into frames and those frames are pre-processed to get the enhanced features. These features are then extracted and saved in a codebook. For testing purpose, the input video is taken, it is pre-processed and features are extracted from it generating a code vector. This code vector is matched with the existing reference codebook and gesture is recognized. After trying variations of multiple algorithms for Pre-processing, Feature extraction and vector quantization, the best performing algorithm was shortlisted to be a combined output algorithm for pre-processing, 2D FFT Fourier Descriptors for feature extraction and 4 vector codebook LBG. The system has been implemented with the above mentioned algorithms to obtain final output. The system is implemented for 10 users and the gestural data of 10 numbers, 26 alphabets and 10 different phrases. The results show overall accuracy of the system to be 92.91% when Numbers, Alphabets and Phrases are considered together. Where the accuracy of Alphabets individually is 85.73%, accuracy of numbers individually is 95.5% and accuracy of phrases is as high as 97.5%.

VII. ACKNOWLEDGEMENT

Authors would like to express their deep gratitude towards Mrs. Kinjal Shah, In-charge of Indian Sign Language Cell, AYJNIHH, Mumbai for her kind support, encouragement and precious time. Also our special thanks goes to all the signers who allowed us to capture the gestures and their multiple iterations without any hesitation and allowed to include their signs in this paper.

VIII. REFERENCES

[1] M. Sushmita, A. Tinku, *"Gesture Recognition: A Survey",* IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 37, issue 3, pp. 311-324 May 2007

[2] N. Purva, K. Vaishali, *"Gesture Recognition: A RevolutionaryTool"*, International Journal of Technological Advancement and Research Volume 3 Issue 3, pp. 24-28, 2013.

[3] N. Purva, K. Vaishali, *"Indian Sign language Recognition: A Review",* IEEE proceedings on International Conference on Electronics and Communication Systems, pp. 452-456, 2014

[4] U. Zeshan, M. Vasishta, S. Meher, *"Implementation of Indian Sign Language in Educational Settings",* Asia Pacific Disability Rehabilitation Journal vol 16, no.1, pp. 16-40, 2005

[5] F. Pravin, D. Rajiv, *""HASTA MUDRA" An Interpretation of Indian Sign Hand Gestures",* 3rd International conference on Electronics Computer technology, vol. 2, pp377-380, 2011

[6] K. Lalit, K. Pritee, *"Towards Establishing a Mute Communication an Indian Sign Language Perspective"*, IEEE Proceedings of 4th International Conference on Intelligent Human Computer Interaction, Kharagpur, India, December, 2012

[7] A. Aleem, M. Yousuf, U. Mehmood, M. Suleman, R. Sameer, U. Razi, A. Israr , *"Pakistan Sign Language Recognition Using Statistical Template Matching",* World Academy of Science, Engineering and Technology International Journal of Computer, Information Science and Engineering ,vol.1, no:3, pp. 286-289, 2007

[8] P. Subha Rajam, G. Balakrishnan, *"Real Time Indian Sign Language Recognition System to aid Deaf-dumb People"*, 13th International conference on Communication technology, pp. 737-742, 2011.

[9] J. Rekha, J. Bhattacharya, S. Majumder, *"Shape, Texture and Local Movement Hand Gesture Features for Indian Sign Language Recognition",* 3rd International conference on Trends in Information Sciences and Computing, pp. 30-35, 2011.

[10] D. Divya, B. Nikesh, *"Indian Sign Language Recognition"*, 2012 1st International Conference on Emerging Technology Trends in Electronics, Communication and Networking, 2012

[11] A. C. Subhash, J. Anand, B. Charul, *"Recognition of Indian Sign Language using Feature Fusion"*, IEEE Proceedings of 4th International Conference on Intelligent Human Computer Interaction, Kharagpur, India, December, 2012

[12] N. Purva, K.Vaishali, *"Indian Sign Language Recognition: Database Creation, Hand Tracking and Segmentation"*, IEEE Proceedings on Circuits, Systems, Communication and Information Technology applications, 2014

[13] WEBSITE: http://ayjnihh.nic.in/index.asp

[14] F.S. Chen, C. Fu, & C. Huang, *"Hand gesture recognition using a real-time tracking method and hidden Markov models"*, ELSEVIER, Image and Vision Computing, Volume 21, Issue 8, pp. 745–758, 2003.

[15] K. Ferens, W. Lehn, and W. Kinsner, *"Image Compression Using Learned Vector Quantization",* in proceedings oh WESCANEX, 1993.

[16] Chin-Chen, H. Yu-Chen, *"A Fast LBG Codebook Training Algorithm For Vector Quantization",* IEEE transactions on consumer Electronics, vol 44, Issue 4, page 1201-1208, 1998