Real Time Sign Language Recognition using the Leap Motion Controller

Deepali Naglot
Research Scholar
Department of CSE (IT),
Vishwakarma Institute of Technology,
Pune, India
dnaglot1711@gmail.com

Milind Kulkarni
Assistant Professor
Department of CSE (IT),
Vishwakarma Institute of Technology,
Pune, India
milind.kulkarni@vit.edu

***Abstract*—Hearing and speech impaired people use Sign Language to convey their message to normal people. Sign Language has evolved as one of the major areas of research and study in computer vision. Researchers in sign language recognition used different input devices such as data gloves, web camera, depth camera, color camera, Microsoft's Kinect sensor, etc. to capture hand signs. In this paper we display the importance of American Sign Language and proposed technique for classification and their efficient results. American Sign Language uses only one hand to display the gestures and thus makes it easy for interpretation and understanding.**

The signs are captured using new digital sensor called "Leap Motion Controller". LMC is 3D non-contact motion sensor which can tracks and detects hands, fingers, bones and finger-like objects. Proposed system used Multi-Layer Perceptron (MLP) neural network with Back Propagation (BP) algorithm to build a classification model which takes feature set as input. Multi-Layer Perceptron (MLP) neural network used to recognize different signs. We have considered 26 different alphabets of American Sign Language. Multi-Layer Perceptron (MLP) is executed on a dataset of total 520 samples (consisting of 20 samples of each alphabet). Recognition rate of proposed system is 96.15%.

Keywords-Leap Motion Controller (LMC), American Sign Language (ASL), Sign Language, Multi-Layer Perceptron (MLP)

I. INTRODUCTION

For more than 250-300 million people around the world, hearing loss and disability to speak presents everyday challenges-some large, some small. To build an application for the deaf and dumb community has become of utmost importance today. With the changing world in terms of technology innovations and education opening new arenas of opportunities, it becomes necessary for everyone to be ahead of the others in this race. In order to enable the deaf and dumb community to create recognition and also to give them a standard platform to communicate and express their opinions with every other individual, this application is being created [13].

A sign language uses hand gestures and body language in order to communicate with the other end. On the other end, it could be either a deaf and dumb person wishing to communicate or a normal person. With an initiative to build a unified interface between the two, this paper focuses at the essential step needed to the system. Proposed system would convert text to American Sign Languages. This would make communication among all people easier and convenient.

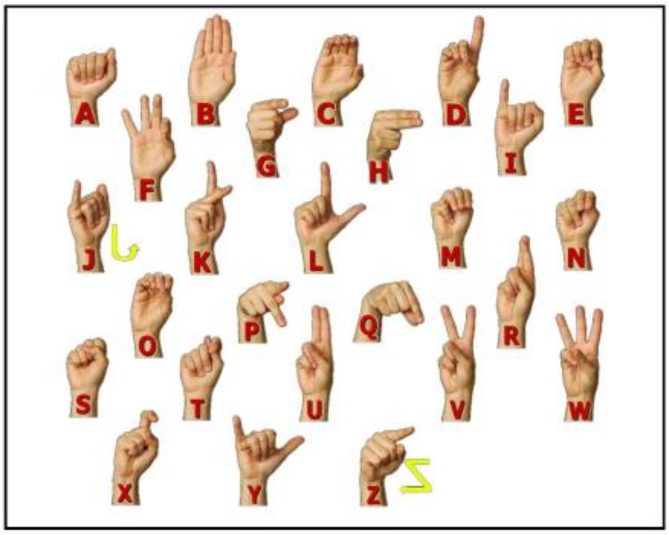

Fig. 1: 26 alphabets of American Sign Language [14]

Many researchers have been working on different sign languages like Arabic Sign Language (ArSL), Australian Sign Language, British Sign Language (BSL), Danish Sign Language French Sign Language, Indian Sign Language (ISL), Taiwan Sign Language etc. In this paper, the fundamental experimentation is done on American Sign Language. American Sign Language, considered to be one of the predominant sign language for deaf communities in United States. ASL is widely used language all throughout the world. It uses only one hand to display the gestures and thus makes it easy for interpretation and understanding. It comprises of around 6000 gestures and other common words. The common words are shown with some specific gesture or spelling with the help of 26 hand gestures indicating 26 alphabets of ASL [13].In this paper, we have considered total of 26 alphabets of ASL which is shown in Fig.1.

The rest of the paper is organized as follows: Section II discusses about the literature survey in the area of sign language recognition using different sensors. The proposed method is presented in detail in section III .Section IV presents experimental results and discusses about performance evaluation parameter. Section V gives the conclusions of the work.

II. LITERATURE SURVEY

Cao Dong, Ming C. Leu and Zhaozheng Yin used Microsoft Kinect sensor to recognize American Sign Language alphabet**.** This paper used Microsoft's Kinect sensor to obtain depth data. The Hand segmentation is done using per-pixel classification method. Random Forest (RF) gesture classifier was implemented to recognize ASL signs using the joint angles. The system considered 24 static alphabets and achieved accuracy of 92% [1].

In [2], Adithya and Vinod presented a method of recognizing the Indian sign language. Images are captured using web camera under lighting and environmental condition. Captured images are then converted to YCbCr colour space. Distance transform and Discrete Fourier Transform (DFT) is used for extract features from image and artificial neural network used for recognizing Indian sign language alphabets. Black background and static images were taken in the experiments, achieving a recognition rate of 91.11%.

In [3], Kai-Yin Fok, Nuwan Ganganath introduced a method for recognizing digits (0-9) of American Sign Language. The proposed system utilizes two Leap Motion sensors for collecting hand states from different viewing angle. Collected data fused using multiple sensors data fusion (MSDF).Key features of proposed system were tip to palm ratio, tip to tip ratio and tip to joint ratio. The recognition was performed using hidden Markov models (HMM). The achieved recognition accuracy using both sensors was 93.14%.

Ching-Hua Chuan and Eric Regina proposed a system for ASL recognition using leap motion controller .This system uses leap motion sensor to collect data from the user. They used k-nearest neighbour and SVM to classify the 26 alphabets of American Sign Language. Average classification rate of k-nearest neighbour and support vector machine was 72.78% and 79.83% respectively [4].

In [5], M. Mohandes and S. Aliyu used Leap motion controller to recognize Arabic sign language. This system were collected 10 samples of each 28 letter using leap motion sensor. Out of 23 features returned by Leap Motion Controller for each frame of data they selected 12 most relevant features for further process. They used Nave Bayes classifier and Multilayer Perceptron (MLP) to classify 28 letters in Arabic sign language. A correct recognition rate of 99.1% achieved using Multilayer Perceptron and 98.3% using Nave Bayes classifier.

An approach for the recognition of Arabic sign language is addressed in [6]. A.S.Elons and Menna Ahmed used Leap motion sensor. This system works with 50 different dynamic signs from Arabic sign language. The signs were collected from 4 different persons, two signs set used for training and other two for testing. The fingers position and distance between the fingers in each frame were features sets provided to system. Proposed system used Artificial Neural Network (ANN) as classifier to recognize the gestures. Recognition accuracy of 88% was achieved.

Priyanka Mekala, Ying Gao introduced method for real-time American Sign Language recognition. A camera sensor was used to capture the video from the user. Image frames were extracted from video. Backgrounds of image frames were subtracted using Running Gaussian average method. The Discrete Fourier Transform (DFT) used for feature extraction. Combinational Neural Networks minimize the size of search space was used as classifier to detect signs. The algorithm achieved 100% recognition rate [11].

III. PROPOSED SYSTEM

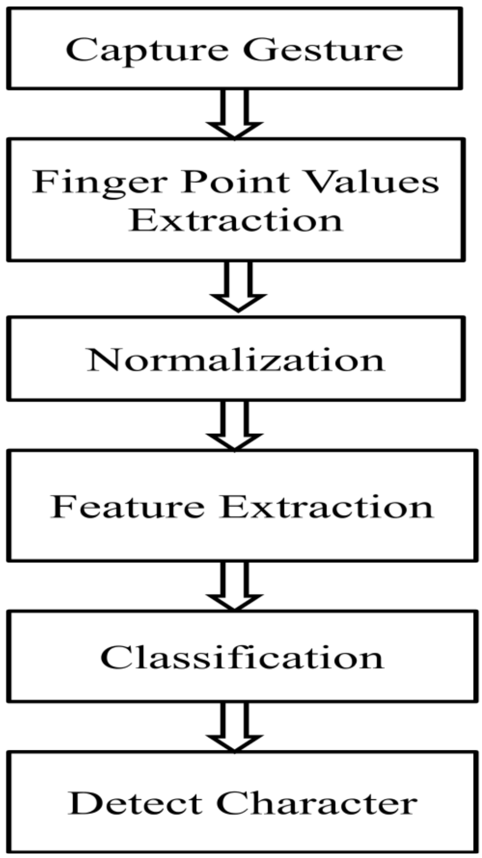

Fig. 2: Architecture of Proposed System

1) Data Collection

The proposed system will take 3D dynamic signs as input. In proposed system, data is collected using leap motion controller**.** The Leap Motion controller is a small USB peripheral device which is designed to allow users to control their computers with hand gestures alone. This sensor is 3D non-contact motion sensor which can detects and tracks hands, fingers, bones and finger-like objects reporting discrete position and motion. The heart of the device consists of two monochromatic IR cameras and three infrared LEDs as shown in Fig. 2. The device has a large interaction space of eight cubic feet and viewing range is approximately 1 inch to 2 feet (60 cm) above the device. The Leap Motion system employs a right-handed Cartesian coordinate system. The origin is centered at the top of the Leap Motion Controller. The x-axis lies in the horizontal plane and running parallel to the long edge of the device. The y-axis lies in vertical plane, with positive values increasing upwards. The z-axis has positive values increasing toward the user and it is lie in the horizontal plane [15, 16].

Dataset of total 520 samples (consisting of 20 samples of each alphabet) from which 260 used for training and remaining samples used for testing. Dataset is normalized before providing to ANN for training and testing.

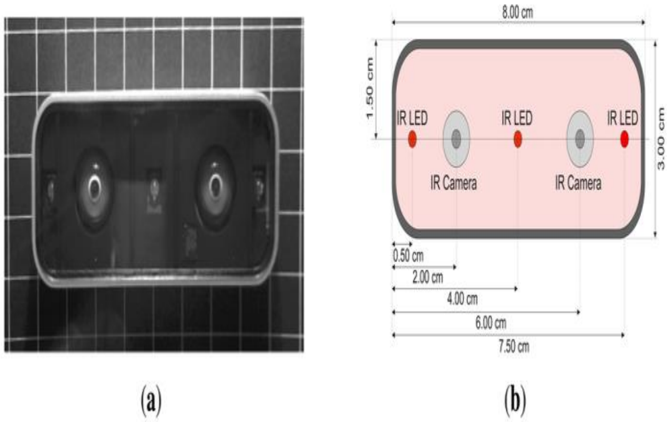

Fig. 3: Leap Motion Controller

2) Features Extraction

Leap motion API include different features for hand, fingers, bones and gesture. Some of them as follow:

Hand: The hand model provides information about the type (left, right or both hand), position (the center position of the palm in millimeters), velocity (in millimeter per second) and other characteristics of a detected hand, the arm to which the hand is attached, and lists of the fingers associated with the hand [16].

Fingers: Fingers related features include finger direction (a unit direction vector), finger length (in millimeter), width, tip position, tip velocity, dip position, pip position, mcp position.

Gestures: Certain movement patterns recognized by leap motion controller. LMC recognizes the motion of a finger tracing a circle in space as a *Circle* gesture, linear movement of a finger as a *Swipe* gesture, downward tapping movement by a finger or tool i.e. tapping a keyboard key as a *Key Tap* gesture, forward tapping movement by a finger or tool i.e. tapping vertical computer screen as a *Screen Tap* gesture [16].

Fig. 4: Features provided by LMC API [16]

From this detailed information, we select the Palm and finger dataset to be the features. Normalized features are provided to ANN. Features used for proposed system to train Artificial Neural Network (ANN) are:-

1) The Euclidian distances between the consecutive finger tip position to palm position.
2) The Euclidian distances between the finger tip position of each consecutive finger.

3) Classification

1) *Multilayer Perceptron neural network (MLP)*

Multilayer Perceptron neural network (MLP) is self-learning, supervised algorithm used for pattern recognition and classification. MLP is feed-forward neural network model that uses feature set as input and transforms it into output units specifying letter/alphabet for given sample. Backpropagation stands for "backward propagation of errors", is a most commonly used algorithm for training neural network for a given set of labelled training examples. The ANN consists of three types of layers input, output and hidden layer with one or multi layer. The input layer neurons are connected to the hidden layer neurons and the hidden layer neurons are connected to output layer neurons by means of interconnection of weights as shown in fig.5. The size input layer is set by number of feature set we want the network to process. Similarly, the size of the output layer is set by the number of alphabets we want to recognize.

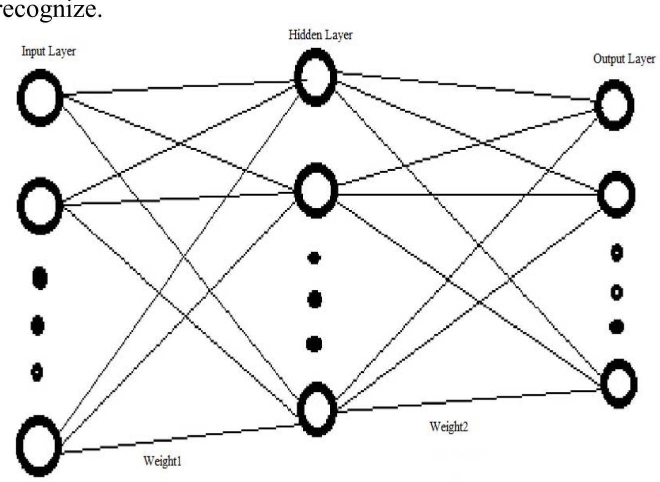

Fig. 5: Architecture of MLP neural network

Networks trained with the Back Propagation (BP) algorithm consist of 3 main steps as following

- Apply input to network and initialize weights and bias
- Forward Pass
  - Calculate net input to the each of the hidden layer unit
  - Calculate output to the each of the hidden layer unit
  - Calculate net input to the each of the output layer unit
  - Calculate net output to the each of the output layer unit
- Reverse Pass
  - Calculate Error
  - Update weights of interconnection from hidden layer unit to output layer unit
  - Calculate Error of hidden layer
  - Update weight of interconnections from input layer neurons to hidden layer neurons

IV. RESULTS AND DISCUSSION

The proposed system was implemented using Intel(R) Core(TM) i5-4210U processor @ 2.40 GHZ speed and the code was written using Java programming language. Data is collected using 4 different persons through Leap motion sensor, total 520 (20 samples of each alphabet) input samples collected. Out of 520 input samples, 260 (10 samples of each alphabet) samples used for training and 260 (10 samples of each alphabet) samples used for testing. Output of the feature extraction step is feature set. It is used as input of the classifier to recognize the sign. Multi-Layer Perceptron (MLP) neural network with Back Propagation (BP) algorithm was used as the classifier and 26 ASL alphabets were classified in the proposed system. Proposed system contains single input layer, hidden layer and output layer. The input layer has 9 neurons which is number of feature set (Euclidian distances between the consecutive finger tip position to palm center and Euclidian distances between the finger tip positions of each consecutive finger), 26 neurons in output layer which is number of alphabets (classes). Number of neuron in hidden layer calculates using (input neuron + output neuron) /2. Sigmoid activation function is used in MLP neural network.

Table 1: Confusion Matrix for 2 class classification

		Predicted Class		
	Class	As Positive	As Negative	
Actual Class	Positive	TP	FN	P
	Negative	FP	TN	N
	Total	P’	N’	P+N

One of the approaches to evaluate performance of the supervised learning algorithms like MLP with BP is confusion matrix sometimes called as contingency table. Confusion matrix is analyzing how well our classifier can recognize samples of different class. A confusion matrix of size k x k associated with a classifier, demonstrates the predicted and actual/target classification, where k is the number of different classes. Table 1.presents confusion matrix, where TP stands for True Positive, FN stands for False Negative, FP stands for True Positive and TN stands for True Negative.

Accuracy is the percentage of test samples that are correctly recognized by the classifier. Precision is measure of exactness. Recall is measure of completeness.

$$\text{Accuracy}=\frac{TP+TN}{P+N} \tag{1}$$

$$\text{Precision}=\frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall}=\frac{TP}{TP+FN} \tag{3}$$

As shown in result, some samples of the alphabets E, S and T were misclassified during testing due to similarity between them. The alphabet N was confused to U, since they are identical to each other. Fig.6 shows the confusion matrix of proposed system with size 26 x 26.

V. CONCLUSION

Sign Language Recognition is essential for the deaf and dumb people to communicate with other people. This paper presents method for American Sign Language recognition using newly introduced Leap Motion Controller. Leap Motion Controller is 3D non-contact motion sensor which can detects and tracks hands, fingers, bones and finger-like objects reporting discrete position and motion. Some of the advantages of Leap Motion Controller are: Robustness, Require less memory, Fast Processing. Proposed system does not require any specific background and environmental condition.

Features provided to ANN to train the system are distance between fingers tip position to palm center and distance between consecutive fingers tip position. Multi-Layer Perceptron (MLP) neural network with Back Propagation (BP) algorithm is used to recognize 26 letters of American Sign Language. Recognition rate of proposed system is 96.15%. The performance of the proposed system is improved in terms of accuracy and time to build the model using above approach. Some gestures are not recognized correctly due to the similarity between few signs as well as orientation of the hand gesture in front of the LMC. Recognition of both hand gestures and sentence or words will be implemented in the future.

Actual Class	Predicted Class																															
Sign	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	Total	TP	FN	FP	Precision	Recall
A	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	0	0	100%	100%
B	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	0	0	100%	100%
C	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	0	0	100%	100%
D	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	0	0	100%	100%
E	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	0	2	83.33%	100%
F	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	0	0	100%	100%
G	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	0	1	90.90%	100%
H	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	0	0	100%	100%
I	0	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	0	0	100%	100%
J	0	0	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	0	0	100%	100%
K	0	0	0	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	0	0	100%	100%
L	0	0	0	0	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	0	0	100%	100%
M	0	0	0	0	0	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	0	0	100%	100%
N	0	0	0	0	0	0	0	0	0	0	0	0	0	**8**	0	0	0	0	0	0	2	0	0	0	0	0	10	8	2	2	80%	80%
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	0	10	10	0	0	100%	100%
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	0	10	10	0	0	100%	100%
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	0	0	10	10	0	0	100%	100%
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**9**	0	0	1	0	0	0	0	0	10	9	1	0	100%	90%
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	10	10	0	2	83.33%	100%
T	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	**6**	0	0	0	0	0	0	10	6	4	0	100%	60%
U	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	**8**	0	0	0	0	0	10	8	2	3	72.72%	80%
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	0	0	0	0	10	10	0	0	100%	100%
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	0	0	0	10	10	0	0	100%	100%
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	0	0	10	10	0	0	100%	100%
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	0	10	10	0	0	100%	100%
Z	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**9**	10	9	1	0	100%	90%
Total	10	10	10	10	12	10	11	10	10	10	10	10	10	10	10	10	10	9	12	6	11	10	10	10	10	9	**260**	**250**	**10**	**10**	**96.54%**	**96.15%**
	Recognition Rate:-**96.15%**																															

Fig 6: Confusion Matrix of Proposed System

VI. REFERENCES

[1] Cao Dong, M. C. Leu and Z. Yin, "American Sign Language alphabet recognition using Microsoft Kinect," 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, 2015, pp. 44-52.

[2] Adithya, V.; Vinod, P.R.; Gopalakrishnan, U., "Artificial neural network based method for Indian sign language recognition," Information & Communication Technologies (ICT), 2013 IEEE Conference on , vol., no., pp.1080,1085,2011-12 April 2013.

[3] K. Y. Fok, N. Ganganath, C. T. Cheng and C. K. Tse, "A Real-Time ASL Recognition System Using Leap Motion Sensors" Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2015 International Conference on, Xi'an, 2015, pp. 411-414.

[4] C. H. Chuan, E. Regina and C. Guardino, "American Sign Language Recognition Using Leap Motion Sensor," Machine Learning and Applications (ICMLA), 2014 13th International Conference on, Detroit, MI, 2014, pp. 541-544.

[5] M. Mohandes, S. Aliyu and M. Deriche, "Arabic sign language recognition using the leap motion controller," 2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE), Istanbul, 2014, pp. 960-965.

[6] A. S. Elons, M. Ahmed, H. Shedid and M. F. Tolba, "Arabic sign language recognition using leap motion sensor," Computer Engineering & Systems (ICCES), 2014 9th International Conference on, Cairo, 2014, pp. 368-373.

[7] Gweth, Y.L.; Plahl, C.; Ney, H., "Enhanced continuous sign language recognition using PCA and neural network features," Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on , vol., no., pp.55,60, 16-21 June 2012.

[8] C. S. Weerasekera, M. H. Jaward and N. Kamrani, "Robust ASL Fingerspelling Recognition Using Local Binary Patterns and Geometric Features," Digital Image Computing: Techniques and Applications (DICTA), 2013 International Conference on, Hobart, TAS, 2013, pp. 1-8.

[9] A. Agarwal and M. K. Thakur, "Sign language recognition using Microsoft Kinect," Contemporary Computing (IC3), 2013 Sixth International Conference on, Noida, 2013, pp. 181-185.

[10] Samrudhi Naik "Depth Sensing Cameras- A Comparative Study, "International Engineering Research Journal (IERJ), Volume 1 Issue 11 Page 1585-1590 ISSN 2395-1621

[11] P. Mekala, Y. Gao, J. Fan and A. Davari, "Real-time sign language recognition based on neural network architecture," 2011 IEEE 43rd Southeastern Symposium on System Theory, Auburn, AL, 2011, pp. 195-199.

[12] Dr.Sudeep Thepade, Arati Narkhede, Priti Kelvekar " Novel Technique for Background Removal from Sign Images for Sign Language Recognition System" International Journal of Computer Applications, (0975 – 8887) Volume 78 – No.10, September 2013.

[13] Dr. Sudeep.D.Thepade Ms.Arati Narkhede, Ms.Priti Kelvekar, Ms.Gandhali Kulkarni and Ms.Seema Tathe, " Sign Language Recognition using Color Means of Gradient Slope Magnitude Edge Images" published at International Conference on Intelligent Systems and Signal Processing (ISSP2013), 978-81-909376-6-5 2013IEEE.

[14] http://lifeprint.com/as1101/topics/wallpaper1.html

[15] A. Colgan, "How Does the Leap Motion Controller Work?",http://blog.leapmotion.com/hardwareto-software-how-does-the-leap-motion-controller-work

[16] Leap Motion, http://www.leapmotion.com