VISION BASED HAND GESTURE RECOGNITION USING FOURIER DESCRIPTOR FOR INDIAN SIGN LANGUAGE

Archana Ghotkar[1], PujashreeVidap[1] and Santosh Ghotkar[2]

[1]Department of Computer Engineering, Pune Institute of Computer Technology, Savitribai Phule Pune University, Pune, India
[2]Department of Computer Engineering and IT, College of Engineering, Savitribai Phule Pune University, Pune, India

ABSTRACT

Indian Sign Language (ISL) interpretation is the major research work going on to aid Indian deaf and dumb people. Considering the limitation of glove/sensor based approach, vision based approach was considered for ISL interpretation system. Among different human modalities, hand is the primarily used modality to any sign language interpretation system so, hand gesture was used for recognition of manual alphabets and numbers. ISL consists of manual alphabets, numbers as well as large set of vocabulary with grammar. In this paper, methodology for recognition of static ISL manual alphabets, number and static symbols is given. ISL alphabet consists of single handed and two handed sign. Fourier descriptor as a feature extraction method was chosen due the property of invariant to rotation, scale and translation. True positive rate was achieved 94.15% using nearest neighbourhood classifier with Euclidean distance where sample data were considered with different illumination changes, different skin color and varying distance from camera to signer position.

KEYWORDS

Fourier descriptor, Indian sign language interpretation, Nearest neighbourhood classifier.

1. INTRODUCTION

Interpretation of sign language is one of the application of gesture recognition and need in current eras to bridge the communication gap between hearing impaired people and normal people [1]. It was estimated that in India, there are over a million people who are profoundly deaf and approximately ten million hard of hearing people [4]. As sign language consists of all the body parts such as hand, lips, eyebrow movement facial expression and emotions. Here focus is given on hand body part as it plays major role in any sign language. In this proposed work, considering the limitation of data glove/ sensor based approach vision based hand gesture recognition system was designed using Indian sign language (ISL). System was considering manual alphabets and numbers for recognition and application interface of text and voice was given for recognized gesture. This is an on-going research work for interpretation of ISL. The proposed algorithm developed for both single as well as two handed static ISL alphabet with a single normal web-camera. To recognize the gestures, system consists of five modules: Image acquisition, Pre-

processing, Segmentation, Feature Extraction and Gesture Recognition with application Interface. Hue, Saturation and value (HSV) color model was used for hand segmentation. Fourier descriptor was used as a feature for hand signs. Recognition was done using nearest neighbourhood classification using Euclidean distance. The rest of the paper is organized such as; section 2 gives literature survey for hand gesture recognition system. Section 3 introduced Data set of ISL. In section 4, methodology is presented. Results and discussion has been done in the section 5. In the section 6 Conclusion and Future work is discussed.

2. LITERATURE SURVEY

The sign language used in India is referred as Indian Sign Language (ISL). Linguistic studies on ISL were started around 1978 and it has been found that ISL is complete natural language having its own morphology, phonology, syntax and grammar [4]. The research of the sign language recognition started in 1982. There are two major approaches for sign language recognition: data glove, vision based. Each approach is having its limitation and advantages, but vision based approaches are more feasible as compared to data glove as user need not to wear cumbersome device like data glove[8] [10]. Sign language interpretation using hand gesture recognition for deaf and dumb people is one of the major applications in HCI. As sign languages differs country wise like spoken language. Within the country also SLs are differs region wise. P. Subha et al. [11] used south Indian sign language for Tamil letters of 32 signs using static hand posture. In this paper each sign was represent in five digit binary code considering up and down position of the five fingers. The sign was assigned to each binary code and then converted into text. Recognition system was able to recognize image with 98.125% accuracy when trained with 320 images and tested with 160 images. Daniel Kelly et al. [3] presented Irish Sign Language recognition. Authors have used gestures features such as hand shape, position and movements were used using size function and principal component analysis with the help of color glove. System achieved 93.8% accuracy for correct sign and 9.1 mis-classification rate. As there are different geometric features for the shape such as angle, contour, length, location. Nguyen Dang et al.[9] have used recognition features as location, angle, velocity and motion patterns for American SL alphabets for unconstrained background and accuracy up to 98% was achieved without using additional marker. Real time hand tracking was done by using Kalman filter and for gesture recognition psuedo-2 dimension hidden markov model was used. Rokade et al. [12] proposed thinning method for hand gesture recognition in which recognition features were, an angle and the distance from the end point. Proposed algorithm was also applicable for complex backgroundand static alphabets of American SL with 92.13% accuracy. Tin Maung et al. [13] explained recognition of subset of Myanmar Alphabet Language using neural network. Here recognition feature was used as an orientation histogram which was sensitive to changes in hand orientation. Accuracy was achieved up to 90 for 33 hand postures with advantageous over lighting change condition. Yang et al.[15] explained Chinese Sign Language recognition based on gray level co-occurrence matrix and multi feature fusion like local and global features with 93.094% accuracy under unconstrained environment with support vector machine approach. Britta Bauer and Karl [2] focused on automatic recognition of German continuous sign language using sub- units rather than models for whole signs will be outlined. It will be advantageous for future reduction of necessary training material. This approach employs totally self-organized subunits called fenone. K-means clustering algorithm was used for the definition. Still there are various researchers are working on various sign languages but there is a major need to work on ISL to help Indian hearing impaired people.

3. ISL DATASET

The manual alphabet set were considered for the proposed work. The dataset consists of single as well as two handed static gestures. Initially system was trained under uniform background with different illumination conditions, different skin color and varying distance from camera. ISL alphabet set is shown in figure 1.Total 780 sampled images were captured to train the system.
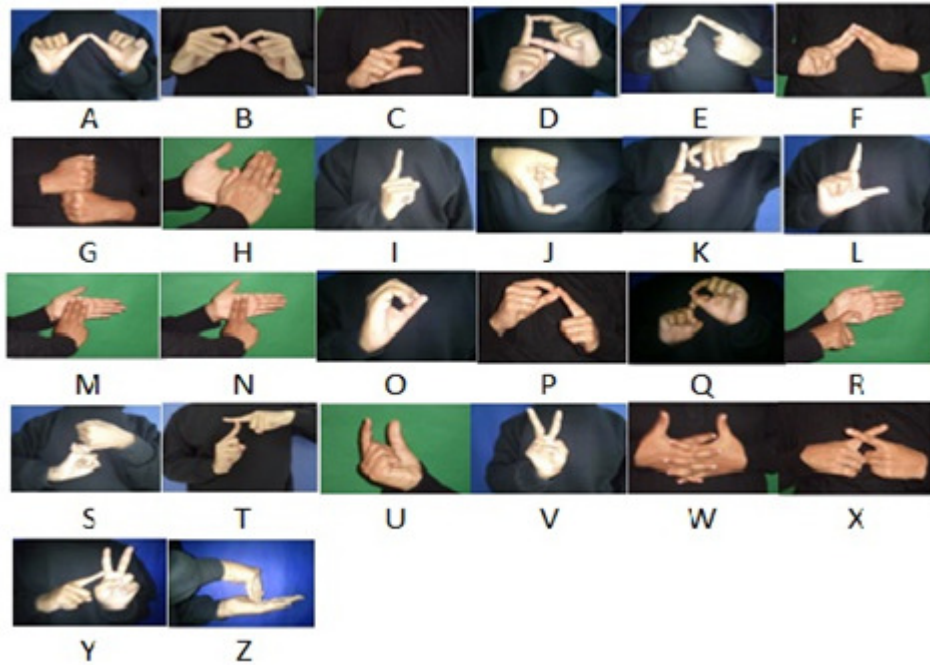

Figure 1.ISL Alphabet Signs [1]

4. METHODOLOGY

It includes modules such as: image acquisition, pre-processing and hand segmentation, feature extraction and recognition. Bare hand segmentation was considered for natural interface. At this stage Fourier descriptor was chosen as a feature vector due to its property of invariant to translation, rotation and scaling. Recognition was done by calculating minimum square root of square error distance using Euclidean distance. Performance analysis was done with nearest neighbourhood (NN) classifier and nearest mean classifier (NMC). For testing, two dataset were considered with variation of hand size and illumination changes and results were analysed. Figure 2 shows the typical flow of any hand gesture recognition system.

4.1. Image Acquisition and Pre-processing

For image capturing web camera of minimum 2 Megapixel under constant background preferably black background was considered to avoid shadow effect. For constant light source night vision web camera was preferred for consistent result. Captured Image was re sized into resolution of 320 X 240 pixels for faster processing. The full sleeves black shirt was used to avoid arm portion of signer. Hand sign training database was created considering different lighting condition and skin color. For captured image low level processing like image blurring using Gaussian blur has

been applied for removal of sharpness to get better segmentation result. Then the image was converted into HSV color space for segmentation [7].
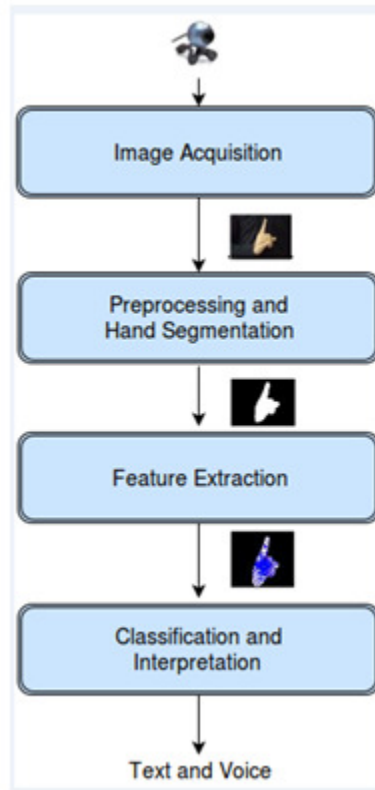

Figure 2. A typical hand gesture recognition system

4.2. Feature Extraction using Fourier Descriptor (FD)[6]

In the proposed work features of the hand signs were calculated using discrete Fourier transform (Fourier descriptor). As signer need to perform the sign in real time, problem of slight variation in angle (rotation), signer hand size(scaling) and translation need to consider while extracting the features. For calculation of features, first segmented image was scanned for finding centroid of the hand blob. The centroid of the hand was calculated by equation 1.

$$x_c = \frac{1}{N}\sum_{i=0}^{N-1} x(i), y_c = \frac{1}{N}\sum_{i=0}^{N-1} y(i) \tag{1}$$

The centroid distance was calculated by distance of boundary points from the centroid (x_c, y_c) and given by equation 2.

$$r(i) = \sqrt{([x(i) - x_c]^2) + ([y(i) - y_c]^2)} \tag{2}$$

The radius of the blob was calculated and boundary pixels were collected for calculation of Fourier descriptor. The 256 Fourier descriptors of the image has been calculated and stored in database. The discrete Fourier transform of r (i) was made. The phase information in the

coefficient was ignored, and the magnitudes of the coefficient were retained which is invariant to translation, scaling, rotation. Fourier descriptor (FD) of this shape is defined in equation 3.

$$[Z_n] = \text{DFT}[r(i)] = \frac{1}{N} \sum_{i=0}^{N-1} r(i) e^{-\frac{2\pi i n}{N}} \qquad (3)$$

$$\text{here, } n = 1 \dots 256, N = 256$$

For each alphabet sign Fourier descriptor was created and stored in a vector form in the database for recognition phase. Algorithm 1 gives detail explanation of feature extraction using Fourier descriptor.

Algorithm 1: Feature extraction using Fourier descriptor

Procedure ExtractFDFeature(SegImage I)

1: Calculate centroid of the binary image
2: $x_c = \frac{1}{N} \sum_{i=0}^{N-1} x(i), y_c = \frac{1}{N} \sum_{i=0}^{N-1} y(i)$// N is number of hand object pixels
3: Calculate shape signature using centroid distance // achieve translation invariance
4: $r(i) = \sqrt{([x(i) - x_c]^2) + ([y(i) - y_c]^2)}$ //calculate $r_0 \dots r_{255}$ using equal angle sampling where x,y ∈ boundary points
5: *For* i= 0 to 255
6: fd_i =FDFT (r(i)) // apply DFT on shape signature r(i)
7: *End For*
8: Get $\vec{fd} = \sum_{i=0}^{255} fd_i$
9: Take magnitude values of fds and ignore phase information // achieve rotation invariance
10: $\vec{fd} = \frac{|fd_1|}{|fd_0|} \dots \frac{|fd_N|}{|fd_0|}$ // achieve scale invariance by dividing DC component
11: $\vec{fd} = (f_{10} \dots f_{255})$ // neglect lower descriptor

EndProcedure

4.3. Recognition

Using similarity distance measures ISL static hand gesture was recognized and displayed the output in the text form (notepad) as well as in the voice. Here, recognition results were analyzed with two approaches.

i) **NMC classifier:**Total 26 mean values of A-Z classes were calculated from training data samples and mean value of each cluster which corresponds with   the   minimum distance with the   unknown gesture declared as the recognized gesture. Euclidean distance measure [5] was used for distance calculation with Fourier descriptor which is a direct measure of the similarity between two multi-dimensional vectors. It is a useful technique for pattern recognition by analysing groups of vector data. The arithmetic mean of feature vector of each training alphabet sign was calculated and stored into mean database file for classification.

$$A = \{a - z\}: u \text{ is training sample}$$

$$\forall u \in A \mid \bar{u} = \frac{1}{n}\sum_{i=1}^{n} \bar{u}_i , n \text{ is number of training samples}$$

Here, the Euclidean distance between two n-dimensional vectors, mean of each alphabet sign {u} and unknown {v} is given as:

$$\bar{u} = (\bar{u}_1, \bar{u}_2, \bar{u}_3, \dots, \bar{u}_n), v = (v_1, v_2, v_3, \dots, v_n)$$

$$D(\bar{u}, v) = \sqrt{\sum_{i=1}^{n} (\bar{u}_i - v_i)^2}$$

ii) **NN classifier:** To compare and analyse the performance with other classifier, analysis has been carried out with other non-parametric classifier such as a nearest neighbourhood with Euclidean distance measures [5].The alphabet from training database, which corresponds with the minimum distance with the unknown gesture,is the recognized gesture. Algorithm 2 explains the steps for NN algorithm.

Algorithm 2: Algorithm for NN

$A = (X_1, \omega_1), (X_2, \omega_2), \dots, (X_n, \omega_n)$// A is the set of N training pattern

//where X_i is of dimension n and ω_i is the class label of the ith pattern

$D(u, x) = \sqrt{\sum_{i=0}^{i=n}(u_i - x_i)^2}$ // Euclidean distance measure

Procedure NN

1. $X_t = \text{TestSign}$
2. for $i \leftarrow 1$, n do   //for n samples
3. $d(X_t, X_i) = \min\{D(X_t, X_i)\}$
4.   end for
5. $X_t = \omega_k$
// pattern X_t is assigned to a class ω_k associated with X_k
End Procedure NN

5. RESULTS AND DISCUSSION

In this paper, ISL alphabets were considered for interpretation using Fourier descriptor. Initial level static black background has been considered while signing. Five persons data set of total 780 signs of A-Z alphabet were considered. Here 65% samples were used for training and 35% samples were used for testing. Here two dataset were considered for testing the algorithm. With NMC approach, Dataset 1 which contained sample images with varying distance, different illumination lighting and variable change in shapes of different persons gave average true positive rate 76%, whereas dataset 2 contained similar shape for signs of different persons with fixed distance achieved up to 79.23% true positive rate. Using nearest neighbourhood, non-parametric

classifier result could increase up to 94.15%. Figure 3 shows comparative analysis of two dataset. Comparison of true positive rate for similarity measure using NMC (mean) and NN is given in Figure 4. Figure 5 shows the confusion matrix of A-Z alphabet where diagonal elements shows true positive rate for unknown samples using NMC approach. The recognition result can be improved by adding more robust features as well as efficient classifier. Considering different illumination condition, different skin colors as well as variable distance of signer hand from camera, sampled images were captured and used for training and testing. Result shows that ideal distance, same illumination condition definitely improved true positive rate. Here, arithmetic mean was considered for similarity measurement to reduce searching complexity but due to the large number of classes true positive rate was degraded. Nearest neighbourhood approach showed improvement in true positive rate but cost of searching complexity increases.
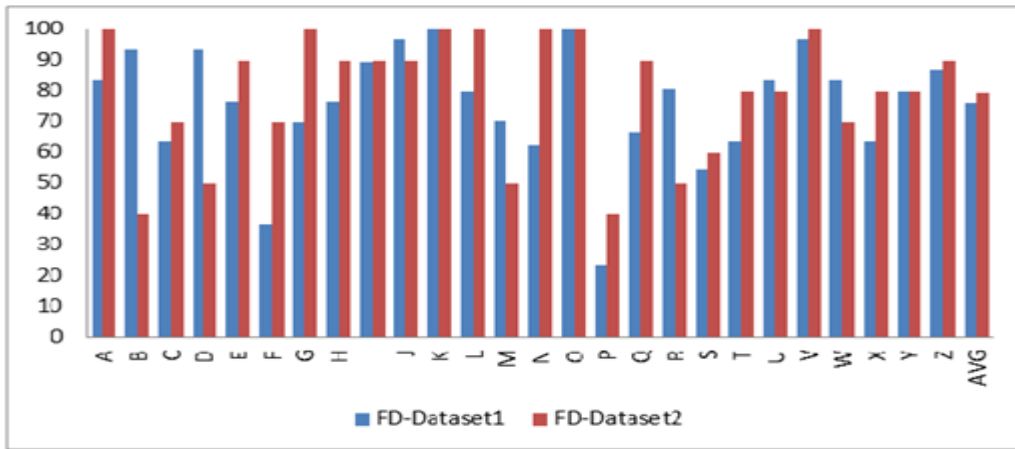

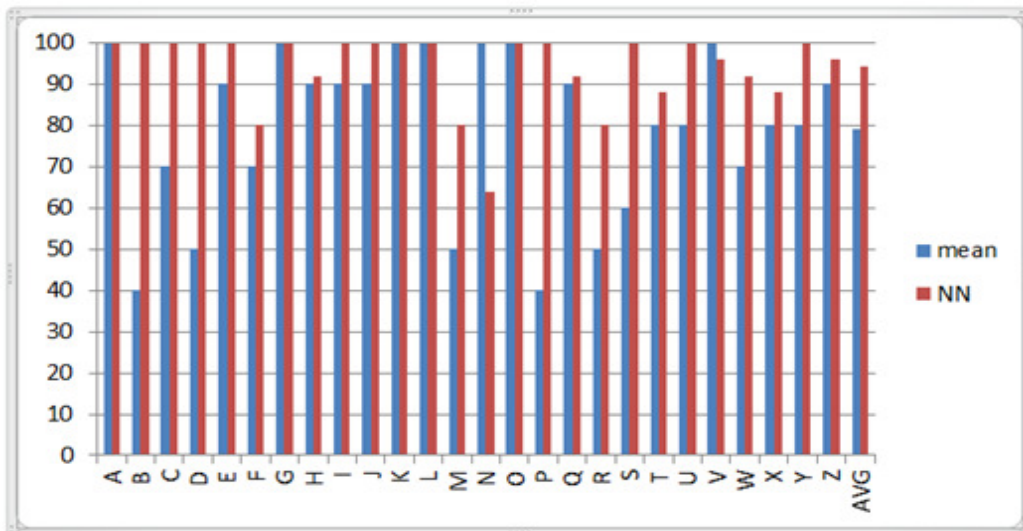Figure 3.Recognition Results of ISL alphabets for two different dataset.

Figure 4.Comparision of true positive rate for alphabets using NMC (mean) and NN

%	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
A	83					6.7										7	3									
B		93									6.7															
C			63					13.3	3.3	17											3.3					
D				93									3.3							3						
E					77	13											7				3					
F	3.3				30	37											10				17				3	
G							70			3.3	10			3	13											
H		3.3	3.3					76.7										13				3.3				
I			3.3					3.3	89														3			
J									3.3	97																
K												100														
L			6.7						6.7			80	6.7													
M							7.1	3.7				4	70	7			7									
N													10	60				10						17		
O															100											
P	30		3.3			6.7										23				7				30		
Q														3			67							8	3	
R							3.8	7.69	3.8				3.8					81								
S										14	18				23				55							
T					10	10											7	10		63						
U			3.3					13.3													83					
V																						97			3	
W							3.3			3.3	3.3								10		3.3		83			
X			3.3	3.3							10			3				3	3	3	3			63	3	
Y		6.7										3.3					7						3		80	
Z														3			8							8	8	87

Figure 5.Confusion matrix for ISL alphabets A-Z using FD with NMC.

6. CONCLUSION AND FUTURE WORK

The hand gesture recognition system was designed for recognition of ISL alphabets and digits as well as static symbols of Indian Sign Language in a common vision based platform. Here, result of ISL manual alphabet is presented and analysed for Fourier descriptor with two classifiers. The system is suitable for both single handed and two handed gestures. To increase the recognition efficiency, contour based as well as region based shape descriptor need to be considered.

Currently, research work is going on 3-D image data which includes depth information [16] for gesture processing with various use of depth cameras such as leap motion controller [17], Kinect sensor [18] and so on. The future work will be concentrate on ISL symbols and sentences which will include the language processing engine with depth sensors. For the whole sign language interpretation incorporation of other body part such as face, lips and emotions are essential but considering limitation of vision based system initially, concentrated on hand modality as it is playing active role in any sign language recognition as well as any human computer interaction applications.

REFERENCES

[1]   Ghotkar, A. S. &Kharate,G. K. (2014) "Study of Vision based Hand Gesture Recognition Using Indian Sign Language", International Journal on Smart Sensing and Intelligent System, Vol. No. 57, No. 1, pp.96-115.

[2]   Bauer, B.&Karl-Friedrich, (2002) "Towards   an   Automatic Sign Language Recognition System Using Subunits",GW 2001,LNAI 2298, pp34- 47,2002,Springer-Verlag Berlin Heidelberg.

[3]   Kelly, D. ,McDonald, Tom Lysaght, Charles, (2008) "Analysis of Sign language Gestures Using Size functions and Principal Component Analysis", International Machine Vision and Image Processing Conference, IEEE Computer Society, pp31-36.

[4]   Dasgupta, Shukla, SandeepKumar, Diwakar,Basu(2008) "A Multilingual Multi- media Indian Sign Language Dictionary Tool", 6th Workshop on Asian Language Resources, 57-64.

[5]   Dong, G. , Jain,A. K.,Ma, W. , Zhang, H. , (2002) "Learning Similarity Measure for Natural Image Retrieval with Relevance Feedback", IEEE Transactions on Neural Networks,Vol.13, No. 4, pp.811-820.

[6]   Zhang, G., Ma, M., Y. He, T. Zhao, (2008) "Shape Feature Extraction Using Fourier Descriptors with Brightness in Content-based Medical Image Retrieval", International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IEEE Computer Society.

[7]   Howe,N. W., Wong, F.& Chekima, A., (2008) "Comparison of Hand Segmentation Methodologies for Hand Gesture Recognition", IEEE-978-4244-2328-6.

[8]   Saliza,N.,Jais,J., (2006) "Hand Gesture Recognition using Hidden Markov Models:A Review on Techniques and Approaches", The Second Malaysian Software Engineering Conference, pp.1-6.

[9]   Binh, N., Shuichi, E., Ejima, T., (2005) "Real time Hand Tracking and Gesture Recognition System", GVIP 05 Conference,Egypt.

[10]   Garg,P.,Aggarwal,   N.&Sofat,   S.   (2009)"Vision   based   Hand   Gesture   Recognition", PWASET,Vol.37,pp.1024-1029.

[11]   Subha, P.& Balkrishnan, G.(2011) "Recognition of Tamil Sign Language Alphabet using Image Processing to aid Deaf-Dumb People", International Conference on Communication Technology and System Design,pp.861-868.

[12]   Rokade,R., Doye&Kokare (2009) "Gesture Recognition by Thinning Method", International Conference on Digital Image Processing, IEEE Computer Society,pp.284-287.

[13]   Maung, T (2009) "Real-Time Hand Tracking and Gesture Recognition System Using Neural Networks",PWASET,Vol.38,pp.470-474.

[14]   Zeshan, U.,Vasishta, M. &Sethna, M. (2005) "Implementation of Indian Sign Language In Educational Setting", Asia pacific Disability Rehabilitation Journal,Vol.16,No.1.

[15]   Quan, Y., Jinye,P. and Yulong,L. (2009) "Chinese Sign Language Recognition Based on Gray Level Co-Occurrence Matrix and Other Multi-features Fusion", IEEE-ICIEA, pp.1569-1572.

[16] Chong, Y., Huang, J. and Pan. (2016) " Hand Gesture Recognition Using Appearance Features Based on 3D Point Cloud", Journal of Software Engineering and Applications, Vol. 9, pp. 103-111.

[17] Naidu   C., Ghotkar A. [2016], "Hand Gesture Recognition Using Leap Motion Controller", International Journal of Science and Research, Vol.5, Issue-10, pp.436-441.

[18] Ghotkar, A., Vidap P. and Deo, K.(2016), "Dynamic Hand Gesture Recognition using Hidden Markov Model by Microsoft Kinect Sensor", International Journal of Computer Application, Vol.150, No.5, pp.5-9.