International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

# Hand Gesture Recognition Using Leap Motion Controller

**Chetna Naidu[1], Archana Ghotkar[2]**

[1]Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India

**Abstract:** ***Sign Language is one of the widely used methods of communication within the hearing impaired people. Sign language is a mode of communication that provides a way of interaction to those hard of hearing, using a collection of gestures and symbols. Hand gesture recognition is a major step for sign language recognition which uses the gestures and symbols to interpret the thoughts of the person. Hand recognition by conventional methods using 2-D cameras suffer from instability due to lighting and skin color variations. The upcoming depth sensors overcome the limitations of the conventional cameras. A Leap Motion controller helps to obtain descriptive information of hand gestures. It tracks the hand and finger movements in digital format and gives few key points associated with each gesture. These key points are used for training and recognition. The paper illustrates the use of Leap Motion controller for hand gesture recognition of Indian sign language signs using four similarity measures namely Euclidean distance measure, Cosine similarity, Jaccard similarity and Dice similarity. 45 gestures of Indian Sign Language including 26 alphabets (a to z),10 numbers(1 to 10) and 9 words are tested with 10 different signers. Average recognition accuracy that is achieved with cosine similarity is 90%, Euclidean distance measure is 88.22%, Jaccard similarity is 86% and with dice similarity is 83.11% using 8 features.***

**Keywords:** Sign language, Hand gesture recognition, Depth sensors, Leap Motion controller.

## 1. Introduction

Human Computer Interaction (HCI) plays an important role in today's world with the increasing applications of computer in everyday life. HCI aims to make human-machine interaction more natural, similar to the human-human interaction [1]. As compared to the other body parts, hand is the most natural and convenient way to express our feelings. Hand gestures are able to represents ideas and actions, using different hand shapes, which can be identified by the gesture recognition system. These are then interpreted to recognize the sign language constructs and provides an interface to the computer system.

For many hearing impaired individuals, sign language is the primary means of interaction. Sign language considers different modes of gestures such as lip movements, facial expression, eye brow movements and hand gestures [1]. These gestures can be classified as manual and non-manual gestures. Manual gestures are the ones that consider only hand movements. Whereas non-manual are the ones that include other features such as facial expressions and lip movements. Gestures can also be classified as static and dynamic gestures. Former are the ones that can be represented in a single image frame. Later are the ones that constitute continuous movement of hands to complete the gesture.

Hand gesture recognition involves sensing techniques for acquiring input from the user. These techniques may be vision based techniques [2] [3], glove based techniques [4] or depth based techniques [12,13]. In glove based techniques, sensors are utilized to measure the joint angles, the position of the hand and fingers in real time [4]. The data glove is dedicated hardware
device for input which is applied to collect the data of joints in order to extract the gesture. This method has the advantages of less input data, high speed. However, gloves are costly and the weights associated with the glove and cables of the measuring equipment restrict the free movement of the hands. This makes the interface complicated for the system.

Vision based techniques uses one or more cameras to capture the gestures performed by the user. In this approach, hand is treated as the direct input equipment. Communication between human and computer needs no other intermediate media. Users can control the machine by making use of the hand gestures. The method of vision-based gesture recognition use camera to collect gesture image sequence, and identifies the gesture by processing and analyze the image [5]. These devices can be used as a translator for people that do not understand sign language, thereby avoiding the intervention of any intermediate person and allows communication using their natural way of speaking. However, these methods are constrained by lighting sensitivity, skin color variations and the distance from the user to the camera. Depth based techniques overcome the limitations of glove and vision based methods by introducing depth sensors. These sensors inspire various hand gesture recognition approaches and applications, which were severely limited in the 2-D domain with conventional cameras.

The depth camera is a distance measuring hardware that is used for 3-D geometric information acquisition. With the advancement in technology of sensor and control devices, there are several devices that yield faster and efficient interaction of humans with computers. The newest sensors provide data that can be successfully used to recognize gestures. The two majorly used depth sensors are kinect sensor and the leap motion controller. Kinectproduces a live stream with depth information, body motion and skeletal movement. Leap motion tracks the user hand movements. Leap Motion, differently from the Kinect directly computes the position of the fingertips of the hand. As compared to other devices, Leap Motion produces a more limited amount of information, only a few keypoints instead of the complete depth map.

Researchers have started exploiting Leap Motion device to develop efficient algorithms that can be used for several applications such as gaming [6, 7],security [8], upper limb rehabilitation [9], human computer interface [10, 11], air-writing [13], etc.

In this paper, Leap Motion controller for hand gesture recognition for static signs of Indian Sign Language (ISL)is used. The system is tested with single handed and double handed gestures of ISL. The Leap Motion controller is a small USB peripheral device which includes two monochromatic IR cameras and three infrared LEDs. The device range is a rough hemispherical area, to a distance of about 1 meter [19]. Leap Motion sensor is a small size sensor which is easy to use and of low cost as shown in figure 1.

**Figure 1:** Leap motion controller [10]

## 2. Literature Survey

Yanmei Chen, Zeyu Ding [1], proposed a rapid early recognition system of dynamic hand gestures based on SVM. The proposed technology of rapid recognition improves the performance of human computer interaction. The dataset consisted of 10 Arabic numbers from 0 to 9 and 26 letters from A to Z, with 100 samples of each gesture and achieved average recognition rate was 82.4%.

D. Bassily and C. Georgoulas [12] in their paper, present the implementation of an intuitive and adaptive manipulation scheme by developing a human-machine communication interface between the Leap Motion controller and a robotic arm. All information regarding the user palm Cartesian position is retrieved from the controller and fed to the algorithm. The algorithm uses the current and previous information supplied by the Leap Controller and achieves an optimum realistic mapping between the user's real palm position and robot's palm position.

ChelsiAgarwal, Debi Dogra [13], present a word extraction and recognition methodology from online cursive handwritten text-lines recorded by Leap motion controller. They propose a segmentation methodology of continuous 3D strokes into text-lines and words by heuristically finding the large gap-information between end and start-positions of successive text lines. Hidden Markov Model-based approach is used to recognize the segmented words and achieved accuracy of 77.6%. Kai-Yin Fok, NuwanGanganath [14], proposed a real-time multisensory recognition system for American sign language (ASL). Data collected from Leap Motion sensors are fused using multiple sensors data fusion (MSDF) and the recognition is performed using hidden Markov models (HMM). The average recognition rate by using the sensors is 93.14%.

Rajesh B. Mapari, GovindKharat [15], developed an Indian Sign Language recognition system is which uses Leap Motion sensor. The recognition of sign posture is done based on Euclidean distance and Cosine similarity. The system was tested for ISL alphabets with 10 different signers. The average recognition accuracy of 88.39% for Euclidean Distance method and 90.32% for Cosine similarity was achieved. M. Mohandes, S. Aliyu [16], propose an approach for Arabic Sign Language Recognition (ArSLR) which involves the use of two Leap Motion Controllers (LMC) to prevent the case of one finger being occluded by another finger or hand. Features are extracted by two controllers and fused together to give 97.7% classification accuracy with Linear Discriminant Analysis (LDA) classifier and 97.1% with classifier level fusion. Better recognition is achieved over the use of a single LMC.

Giulio Marin, Fabio Dominio [17], proposed a hand gesture recognition scheme targeted to Leap Motion and kinect sensors. An ad-hoc feature set based on the positions and orientation of the fingertips is computed and fed into a multi-class SVM classifier in order to recognize the performed gestures. By combining the two features sets, it is possible to achieve a very high accuracy in real-time. The features extracted are distance, fingertip angle and hand orientation and achieved accuracy was 91.28%. A. S. Elons, Menna Ahmed [18], proposed a recognition system for Arabic Sign Language recognition using a Leap Motion Sensor. The temporal and spatial features are fed into a Multi-layer perceptron Neural Network (MLP). The system was tested on 50 different dynamic signs and the recognition accuracy reached 88% for two different persons.

## 3. Dataset

The dataset consists of 45 gestures from Indian Sign Language. It includes 26 alphabets (A to Z), 10 digits (1 to 10) and 9 static words namely, Boss, Bird, Camera, Friend, Enemy, Home, Wrong, Chain, Good.Figure 2 and 3 shows the ISL alphabets and numbers respectively.

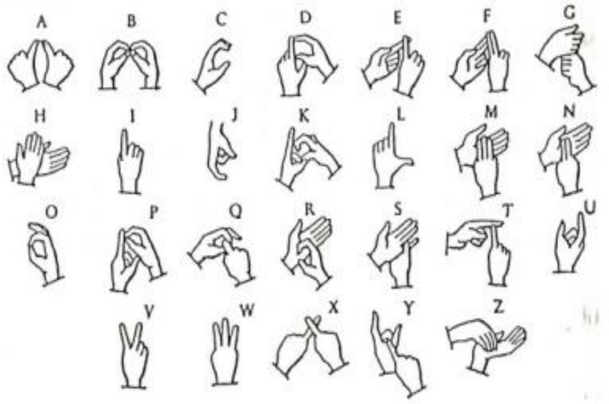

**Figure 2:** ISL alphabets [15]

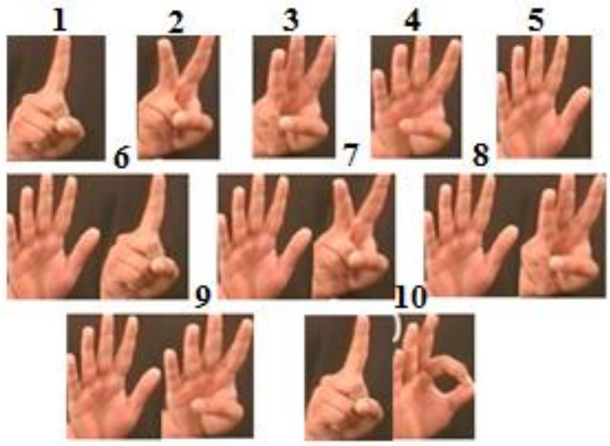

**Figure 3:** ISL numbers [15]

## 4. Methodology

In the proposed system, an input gesture is acquired using a Leap Motion sensor. It tracks hands and fingers movements in 3-D digital format. It gives a mapping of the gesture in terms of feature points. The extracted fingertip positions of each gesture are stored in the database. The distance between them is used as feature vector. While testing, again gestures are captured and the positions of the finger points are extracted. Distances are calculated by Euclidean distance method for current gesture and all gestures stored in the database. For recognition, the feature vectors are compared using the four similarity measures viz. Euclidean distance measure, Cosine similarity, Jaccard Similarity and dice similarity. The one with maximum similarity is returned as the detected gesture. Figure 4gives a detailed view of the architecture of the system.

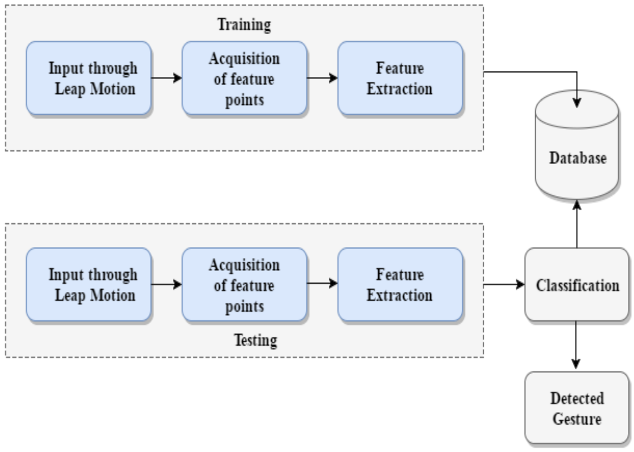


**Figure 4:** Architecture of Proposed System

### A. Data Acquisition

The Leap Motion yields a series of frame by tracking motion of hands and fingers within its view field. The frame data represents a set of tracking data for hands and fingers detected in a single frame. The acquired data through the Leap sensor consists of array of objects to store the physical characteristics of a detected finger such as fingertips. The gestures captured by the Leap controller yields a few keypoints corresponding to each hand gesture.

### B. Feature Extraction

Leap Motion returns a set of relevant hand points and some hand pose features. Figure 5 highlights the features extracted for recognition. The extracted keypoints are coordinates of finger positions from the input gesture. The points are, the center of the palm(say A), tips of thumb(say B), index finger(say C), middle finger(say D), ring finger(say E), pinky finger(say F) for each hand. The coordinates be, A($x_1$, $y_1$, $z_1$), B($x_2$, $y_2$, $z_2$), C($x_3$, $y_3$, $z_3$), D($x_4$, $y_4$, $z_4$), E($x_5$, $y_5$, $z_5$), F($x_6$, $y_6$, $z_6$) for each hand.The extracted points are stored in a database file.

### C. Gesture Classification

Feature points corresponding to each gesture are stored in the database file. During testing, a current gesture is captured and key points are extracted from that gesture. At the run time, distances are calculated from the extracted feature points using Euclidean distance formula as follows:

$$d_i = \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 + (z_i - z_{i+1})^2} \quad (1)$$

Where i= 1 to 8 for single handed gestures and i= 1 to 16 for double handed gestures

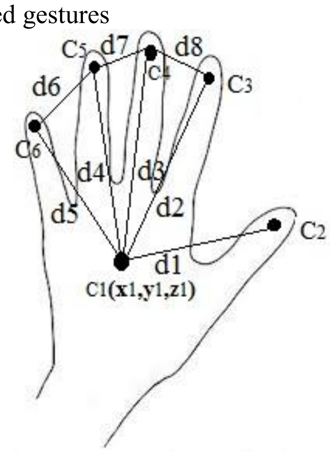


**Figure 5:** Extracted features

As shown in figure 5, $d_1$ gives the distance between points $C_1$ and $C_2$, $d_2$ gives the distance between $C_1$ and $C_3$, $d_3$ gives the distance between points $C_1$ and $C_4$, $d_4$ gives the distance between $C_1$ and $C_5$, $d_5$ gives the distance between points $C_1$ and $C_6$, $d_6$ gives the distance between points $C_5$ and $C_6$, $d_7$ gives the distance between points $C_4$ and $C_5$, $d_8$ gives the distance between points $C_3$ and $C_4$.

- $d_1 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$
- $d_2 = \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2 + (z_1 - z_3)^2}$
- $d_3 = \sqrt{(x_1 - x_4)^2 + (y_1 - y_4)^2 + (z_1 - z_4)^2}$
- $d_4 = \sqrt{(x_1 - x_5)^2 + (y_1 - y_5)^2 + (z_1 - z_5)^2}$
- $d_5 = \sqrt{(x_1 - x_6)^2 + (y_1 - y_6)^2 + (z_1 - z_6)^2}$
- $d_6 = \sqrt{(x_3 - x_4)^2 + (y_3 - y_4)^2 + (z_3 - z_4)^2}$
- $d_7 = \sqrt{(x_4 - x_5)^2 + (y_4 - y_5)^2 + (z_4 - z_5)^2}$
- $d_8 = \sqrt{(x_5 - x_6)^2 + (y_5 - y_6)^2 + (z_5 - z_6)^2}$

Distances are calculated between the points and feature vector is formed. Consider $C_f$ be the vector associated with the current gesture.

$C_f = \{d_1', d_2', d_3', d_4', d_5', d_6', d_7', d_8'\}$ for one handed gestures.
$C_f = \{d_1', d_2', d_3', d_4', d_5', d_6', d_7', d_8', d_9', d_{10}', d_{11}', d_{12}', d_{13}', d_{14}', d_{15}', d_{16}'\}$ for double handed gestures.
For every gesture stored in the database, distances are calculated between the feature points at the runtime.

Consider $C_d$ be the vector.
$C_d = \{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8\}$ for one handed gestures.
$C_d = \{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}, d_{11}, d_{12}, d_{13}, d_{14}, d_{15}, d_{16}\}$ for double handed gestures.
The vector $C_f$ is compared with vector $C_d$ of every gesture in the database and a score is returned. The gesture whose $C_d$ on comparison with $C_f$ returns maximum score, is returned as the detected gesture.
Comparison is done using the following equations:

Euclidean:

$$S_E = \sqrt{(d_1 - d'_1)^2 + (d_2 - d'_2)^2 \ldots + (d_{8/16} - d'_{8/16})^2} \quad (2)$$

Cosine:

$$S_C = \frac{C_f . C_d}{|C_f||C_d|} \quad (3)$$

Jaccard:

$$S_J = \frac{|C_f \cap C_d|}{|C_f \cup C_d|} = \frac{|C_f \cap C_d|}{|C_f| + |C_d| - |C_f \cap C_d|} \quad (4)$$

Dice:

$$S_D = \frac{2|C_f \cap C_d|}{|C_f| + |C_d|} \quad (5)$$

Where $S_E$ $S_C$, $S_J$ and $S_D$ are the scores by Euclidean, Cosine, Jaccard and Dice similarity measures respectively.
The detected gesture is given as the output and results are compared.

## 5. Experimental Results

The system is tested for 45 gestures of ISL, 26 alphabets (A to Z), 10 digits (1 to 10) and 9 static words namely, Boss, Bird, Camera, Friend, Enemy, Home, Wrong, Chain and Good. The experimental results obtained are given in the following tables. Table 1, 2 and 3 give the recognition accuracy for alphabets, numbers and words respectively with four similarity measures. Algorithm is tested on 10 different persons to calculate average accuracy.

**Table 1:** Recognition accuracy for alphabets

| *Method/Gesture* | *Cosine* | *Euclidean* | *Dice* | *Jaccard* |
|---|---|---|---|---|
| A | 90% | 90% | 90% | 90% |
| B | 90% | 90% | 80% | 80% |
| C | 100% | 100% | 100% | 100% |
| D | 80% | 80% | 80% | 80% |
| E | 80% | 80% | 80% | 80% |
| F | 80% | 80% | 80% | 80% |
| G | 80% | 70% | 70% | 70% |
| H | 80% | 70% | 70% | 70% |
| I | 100% | 100% | 100% | 100% |
| J | 90% | 80% | 80% | 80% |
| K | 80% | 80% | 80% | 80% |
| L | 100% | 100% | 100% | 100% |
| M | 90% | 90% | 80% | 70% |
| N | 90% | 80% | 80% | 70% |
| O | 90% | 90% | 90% | 90% |
| P | 80% | 80% | 80% | 80% |
| Q | 80% | 80% | 80% | 80% |
| R | 80% | 80% | 70% | 70% |
| S | 80% | 80% | 70% | 70% |
| T | 100% | 100% | 100% | 100% |
| U | 100% | 100% | 100% | 90% |
| V | 100% | 100% | 100% | 100% |
| W | 90% | 90% | 90% | 80% |
| X | 100% | 100% | 100% | 90% |
| Y | 90% | 80% | 80% | 80% |
| Z | 80% | 80% | 80% | 80% |
| **Average** | **88.46%** | **86.92%** | **85%** | **82.69%** |

**Table 2:** Recognition accuracy for numbers

| Method/Gesture | Cosine | Euclidean | Dice | Jaccard |
|---|---|---|---|---|
| 1 | 100% | 100% | 100% | 100% |
| 2 | 80% | 80% | 80% | 80% |
| 3 | 90% | 90% | 90% | 80% |
| 4 | 100% | 100% | 90% | 80% |
| 5 | 100% | 100% | 100% | 100% |
| 6 | 90% | 90% | 90% | 80% |
| 7 | 90% | 80% | 80% | 80% |
| 8 | 90% | 80% | 80% | 80% |
| 9 | 100% | 100% | 90% | 90% |
| 10 | 90% | 90% | 90% | 80% |
| **Average** | **93%** | **91%** | **89%** | **85%** |

**Table 3:** Recognition accuracy for words

| *Method/Gesture* | *Cosine* | *Euclidean* | *Dice* | *Jaccard* |
|---|---|---|---|---|
| Boss | 90% | 90% | 80% | 80% |
| Bird | 100% | 100% | 100% | 100% |
| Good | 100% | 100% | 100% | 100% |
| Friend | 90% | 90% | 80% | 70% |
| Enemy | 90% | 90% | 80% | 80% |
| Camera | 80% | 70% | 70% | 70% |
| Home | 90% | 90% | 90% | 80% |
| Wrong | 100% | 100% | 100% | 90% |
| Chain | 80% | 70% | 70% | 70% |
| **Average** | **91.11%** | **88.80%** | **85.50%** | **82.22%** |

Following graphs illustrate the results obtained:

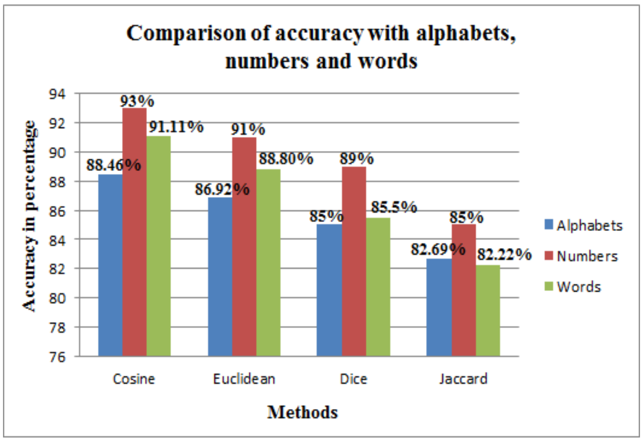


**Figure 6:** Comparison of accuracy with alphabets, numbers and words

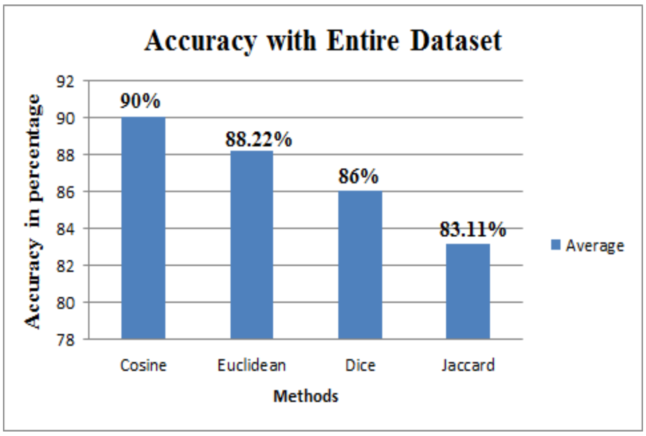


**Figure 7:** Accuracy with complete dataset

Cosine similarity method gives the maximum average accuracy, 88.46% for alphabets, 93.00% for numbers, 91.10% for words and 90.00% for the complete dataset. Euclidean distance method gives average accuracy of 86.92% for alphabets, 91.00% for numbers, 88.80% for words and 88.22% for the complete dataset. Dice similarity gives average accuracy of 85.00% for alphabets, 89.00% for numbers, 85.50% for words and 86.00% for the complete dataset. Jaccard similarity gives average accuracy of 82.69% for alphabets, 85.00% for numbers, 82.22% for words and 83.11% for the complete dataset.

### Comparative Analysis

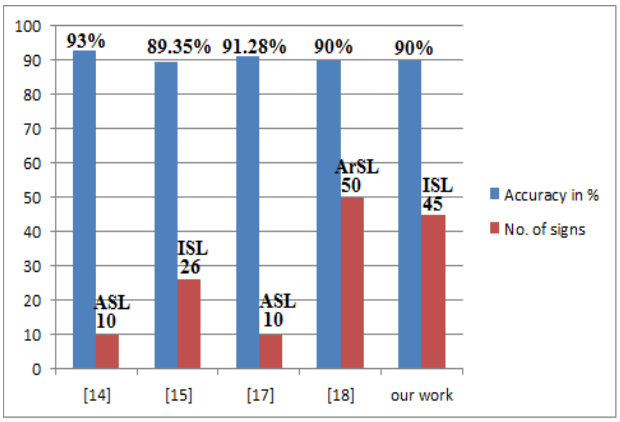


**Figure 8:** Comparative analysis of results

The figure 8,shows the results obtained by various authors for different sign language and different number of gestures. The other authors have achieved accuracy of above 90% for a dataset of with less number of signs. With the increase in number of signs, the recognition accuracy decreases. In the proposed work, for 45 signs of ISL, maximum recognition accuracy of 90% is achieved by using cosine similarity.

## 6. Conclusion and Future Work

The paper presents a scheme to recognize static gestures from ISL using a Leap motion controller. Acquisition of input frame through a leap motion device helps to meet the requirements of efficient gesture recognition in a 3-D space. The sensor tackles the major issues in earlier vision based systems such as skin color, lighting variations and hand orientation relative to the device.

In this work, distance features are used for training and classification. The system is tested using four similarity measures namely Euclidean Distance measure, Cosine similarity, Dice similarity and Jaccard similarity. Analysis of these four distance measure shows that cosine distance gives maximum accuracy of 90%.

The accuracy can be increased by using more features such as finger angles and hand orientation with classifiers such as Bayes and HMM for dynamic gestures. The Leap Motion controller can be integrated with kinect camera to achieve recognition of all body gestures for different sign languages.

## References

[1] Yanmei Chen, Zeyu Ding, Yen-Lun Chen and Xinyu Wu, "Rapid recognition of dynamic hand gestures using Leap Motion", *International Conference on Information and Automation*, pp. 1419-1424, 2015.

[2] S. PadamPriyal andPrabin Kumar Bora, "A robust static hand gesture recognition system using geometry based normalization and krawtchouk moments", *Pattern Recognition*, Volume:46, No:8, pp. 2202-2219, 2013.

[3] Tamer ShanablehandKhaledAssaleh, "User independent recognition of Arabic Sign Language for facilitating communication with the deaf community", *Digital Signal Processing*, Volume: *21*, No:4, pp. 535-542, 2011.

[4] Sun Li-juan, Zhang Li-cai, GuoCai-long, "Technologies of Hand Gesture Recognition Based on Vision", *Computer Technology and Development*, Volume:18, No:10, pp. 214-216, 2008.

[5] WengHan-liang, Zhan Yin-wei, "Vision-Based Hand Gesture Recognition with Multiple Cues", *Computer Engineering and science*, Volume:2, pp. 028, 2012.

[6] M. Rahman, M. Ahmed, A. Qamar, "Modeling therapy rehabilitation sessions using non-invasive serious games", *IEEE International Symposium onMedical Measurements and Applications (MeMeA)*, pp. 1-4, 2014.

[7] Ok-Hue Cho and Sung-Tae Lee, "A study about honey bee dance serious game for kids using hand gesture", *International Journal of Multimedia and Ubiquitous Engineering*, Volume:9, No: 6, pp. 397-404, 2014.

[8] Ishan Nigam , MayankVatsa and Richa Singh, “Leap signature recognition using hoof and hot features”, *International Conference onImage Processing,*pp. 5012-5016, 2014.
[9] D. Charles, K. Pedlow and T. Charles, “An evaluation of the leap motion depth sensing camera for tracking hand and fingers motion in physical therapy”, *Proceedings of the Interactive Technologies and Games Conference,* pp. 028-032, 2013.
[10] Yingying She, Qian Wang, YunzheJia, Ting Gu , Qun He and Baorong Yang, “A real-time hand gesture recognition approach based on motion features of feature points”, *International Conference onComputational Science and Engineering, pp.* 1096-1102*,* 2014.
[11] Qianqian Wang and Yuan-RongXu, “Dynamic Gesture Recognition using 3-D Trajectory”, International Conference on Information Science and Technology (ICIST), pp. 598-601, 2014.
[12] D. Bassily, C. Georgoulas and J. Guettler “Intuitive and Adaptive Robotic Arm Manipulation using the Leap Motion Controller”, *International Symposium on Robotics*, pp. 1-7, 2014.
[13] ChelsiAgarwal, Debi Dogra, RajkumarSaini and ParthaPratim Roy, “ Segmentation and Recognition of Text Written in 3D using Leap Motion Interface”, *Asian Conference on Pattern Recognition(ACPR)*, pp. 539-543, 2015.
[14] Kai-Yin Fok, NuwanGanganath, Chi-Tsun Cheng, and K. Tse Chi, “A Real Time ASL Recognition System using Leap Motion Sensors”, *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, pp. 411-414, 2015.
[15] Rajesh Mapari and GovindKharat, “Real Time Human Pose Recognition Using Leap Motion sensor”, *International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, pp. 323-328, 2015.
[16] M. Mohandes, S. Aliyu and M. Deriche “Prototype Arabic Sign Language Recognition using Multi-Sensor Data Fusion of Two Leap Motion Controllers”, *International Multi-Conference on Systems, Signals & Devices*, pp. 1-6, 2015.
[17] Giulio Marin, Fabio Dominio and PietroZanuttigh, “Hand Gesture Recognition With Leap Motion And Kinect Devices”, *International Conference on Image Processing (ICIP*), pp. 1565-1569, 2014.
[18] A.S.Elons, Menna Ahmed and HwaidaaShedid, “Arabic Sign Language Recognition Using Leap Motion Sensor”, *International Conference on Computer Engineering Systems (ICCES)*, pp. 368-373, 2014.
[19] https://www.leapmotion.com/
[20] http://indiansignlanguage.org/