Virtual Personal Trainer using Microsoft Kinect and Machine Learning

Rashmi A. Rane
Professor
Dept. of Computer Engineering
Maharashtra Institute of Technology, Kothrud,
Pune-411038

Neel Potnis
Dept. of Computer Engineering
Maharashtra Institute of Technology, Kothrud,
Pune-411038

Shrawani Sansare
Dept. of Computer Engineering
Maharashtra Institute of Technology, Kothrud,
Pune-411038

Neekait Mokashi
Dept. of Computer Engineering
Maharashtra Institute of Technology, Kothrud,
Pune-411038.

Sumit Patil
Dept. of Computer Engineering
Maharashtra Institute of Technology, Kothrud,
Pune-411038.

ABSTRACT

Human-Computer Interaction is a flourishing area in terms of research and has many real-world applications. Keeping this in mind, we came up with an idea to develop Human-Computer interaction for proper conduct of physical exercises at home, using information sensed by an RGB-D camera, namely the Microsoft Kinect. Along with Kinect, we also make use of a Machine Learning technique to perform operations on captured data to predict the accuracy of a performed physical exercise. Our approach is based on the study of the movement of various joints in the human body, which we examine with the use of the Kinect. We take into account an algorithm for our implementation - Hidden Markov Model (HMM). We combine these and detect the posture of a user while he performs a particular exercise, before comparing it with our ideal database of postures. Based on this comparison, we predict the accuracy of the exercise and aim to improve and correct the form of the user in terms of performance of the exercise.

General Terms

Microsoft Kinect, Machine Learning, Hidden Markov Model, Unit Vector Generation

Keywords

Physical exercises, exercise accuracy prediction

1. INTRODUCTION

The paper presents a novel technique to perform physical exercise recognition and accuracy prediction by means of an unobtrusive motion sensor device. In particular, the paper particularly adopts Microsoft Kinect as a motion sensor due to its reliability, competitive cost and its usage of tracking. The output of the framework proposed here is basically an accuracy percentage of user performed exercise with respect to ideal physical exercise. The system treats the gestures as a sequence of frames, sampled at regular intervals. This sequence of frames is fed into the system as an input and the system extracts features from each frame, and then after performing computations, the system outputs the accuracy. The system first requires capturing the joint configuration data of a user performing a physical exercise. Then, the system makes the use of unit vector concept, which means that the system generates unit vectors for gesture with the help of a module explained in later chapters. Then, using these unit vectors, the system generates a probability matrix for the gesture using the machine learning technique, namely HMM (Hidden Markov Model). The probability matrix (output of HMM) is then used for the purpose of accuracy prediction which is the desired output. Our current work includes two contributions. The first contribution involves capturing the user performed exercise using Kinect and then predict the accuracy of user performed exercise with respect to ideal physical exercise. The second contribution involves developing on own ideal physical exercise dataset for the purpose of accuracy prediction, for which probability matrices would be computed in advance, and would be used later. The organization of the paper is as follows. Related work is outlined in Section 2. The system architecture and workflow is described in Section 3. Conclusions are presented in Section 4.

2. RELATED WORK

Throughout the research conducted, we came across many works related to HCI using Kinect sensor. The paper for Human activity recognition [1] by Salvatore Gaglio, Giuseppe Lo Re and Marco Morana aims at presenting a method for recognizing human activities using information sensed by Microsoft Kinect. The approach is based on estimation of some relevant joints of the human body by means of Kinect. The data is then operated upon by 3 different machine learning techniques namely K-Means clustering, Support Vector Machines (SVM) and Hidden Markov Model (HMM). These 3 techniques are combined together to detect postures involved while performing a particular activity, to classify them and to model each activity as spatiotemporal evolution of known postures. There are various other works using various approaches and various machine learning techniques. Lorenzo Patras, Ion Giosan, Sergiu Nedevschi's paper [2] presents a system capable of identifying and validating various human body gestures. Body data us acquired form kinect sensor and consist in a set of bones represented by their rotation in 3-D space. The use if Dynamic Time Warping is proposed in the paper for synchronizing the performed gesture with the corresponding ground truth dataset. If the sequences are not synchronized, feedback is returned to the user as a comparison between its performance and the closest sample in the database. The system also consists of a component which records the initial and final position of the user. Body data is captured and obtained as bone rotation. The bone rotations are used for operation by converting the bone rotation into rotation matrices or quaternions. The rotations are obtained as a relative to parent bone. Throughout out research, we also came up to a point that other than Kinect, there can be various devices that can be used to develop such HCI systems. Thomas Schlomer, Benjamin Poppinga, Niels Henze, Susanne Bol's paper [3] proposes to develop a gesture recognition system using a Wii sensor and a controller. The system makes use of a Wii remote controller as an input device along with the Wii-motes acceleration sensor to capture the data. The authors have developed a new library to exploit the Wii sensor data and then make use of HMM model to train the data and recognize the user performed gestures. The Wii sensor is known to produce a constant vector data and so the authors have proposed the need for the usage of vector quantization techniques. Here, they have made use of a simple k-means algorithm to reduce the amount of data they have to deal with and then use the clustered data for training the HMM and the recognizing the user performed gestures. The usage of HMM is such that an HMM is initialized for each gesture and then the famous Baum-Welch algorithm is used for optimizing it. The system proposes the use 2 stage filtering for data preprocessing. Before the recognition process, the vector data is supplied to first filter wherein all the irrelevant vectors are discarded. Then the output from first filter is sent to second filter eliminates all vectors which are roughly equivalent to their predecessor and thus contribute to the characteristic of a gesture only weakly. Some works done in this field [4] make use of unique methods which end up giving a wide range of ideas to be researched upon. Geetha M, Manjusha C, Unnikrishnan P, Harikrishnan R's paper [4] proposes a work to recognize 3-D dynamic signs corresponding to Indian Sign Language words. The work aims to develop a system to capture 3-D dynamic signs of Indian Sign Language using Kinect camera and then perform feature extraction methods to perform gesture recognition. The system makes use of a new trajectory based feature extraction method using the concept of Axis of Least Inertia (ALI). An Eigen distance based method using the seven 3D key points, extracted using Kinect sensor is proposed for local feature extraction. This is method has been proved to improve the performance of the system. After the data is captured form Kinect, B-Spline trajectory is plotted for centroid, index, thumb, middle, ring and small finger motion. These are then subject to feature extraction that extracts the local and global features. The global feature extraction makes use the of the ALI method and the local feature extraction computes the distance between each finger tip to the centroid, in each frame and a matrix is created with all these distance values. This is then subject to Principal Component Analysis (PCA). An integration of both these results helps in identification of gesture correctly. There are also some works that focus upon determining as to which machine learning techniques are better for developing dynamic HCI systems. Alina Delia Calin's paper [5] is a work to analyze the variation of gesture recognition accuracy of several time series classifiers, using the two available versions of Kinect i.e., the older version of kinect sensor which is Kinect 1, and the newer version of Kinect which is Kinect 2. For analysis of several time series gestures, a large database of ideal gestures is developed. The similarity between the user performed and the ideal gestures is analyzed by using 2 machine learning algorithms which are Dynamic Time Warping (DTW) and Hidden Markov Model. The main aim is to first infer as to which version of the kinect sensor is better, kinect 1 or kinect 2. Secondly, the aim is to infer which of the 2 machine learning algorithms are efficient for developing gesture recognition systems, and also which algorithm tends to provide more accuracy in recognition of user gestures. There were no significant difference in classification accuracy between the results obtained by the 2 sensors; however, it was found that Kinect 2 performed better than Kinect 1 in many ways. Her work ends up concluding that HMM is a better option over DTW in terms of accuracy, which makes it preferable to be used in a dynamic system. Marcos Y.O Camada, Jes J.F Cerqueira, Antonio Narcus N. Lim's paper [6] focuses on facilitating the recognition of stereotyped behaviors usually seen during autism. This paper studies the performance between two machine learning algorithms to recognition the Stereotyped gestures typical of autism: (i) Hidden Markov Model [HMM]; and (ii) Support Vector Machine [SVM]. Sequence of orientations data from some joints obtained through a RGB-D (Red Green Blue -Depth) camera [Kinect] are used for analysis. The results of these two machine learning algorithms are compared with state-of-the-art. The HMM approach proposed in this paper have shown 98.89% average recognition rate and 98.9% recall. The works mentioned so far involve no external object detection. The paper [7] focuses on based human body detection and pose estimation along with incorporating attached props. This work proposes a novel method for generating training data of human postures with attached objects. The results have shown a significant increase in body-part classification accuracy for subjects with props from 60% to 94% using the generated image set. This work contributes in extending the Buys framework to accommodate attached props. The modified framework generates virtualized training depth and labeled images for all body parts and an external object. this problem, a real pre-captured motion data will be used to animate a virtual human model. By doing this, different activities can be mapped to many human models with various mesh topologies. After a detailed research on aforementioned as well as a few other works like in [8], [9], [10], [11], we decided to go with Kinect as our hardware and HMM as a machine learning technique.

3. VIRTUAL PERSONAL TRAINER SYSTEM ARCHITECTURE

The entire system of Virtual Personal Trainer is divided into three modules along with an external device and a database. The system architecture is as shown in Figure 1.

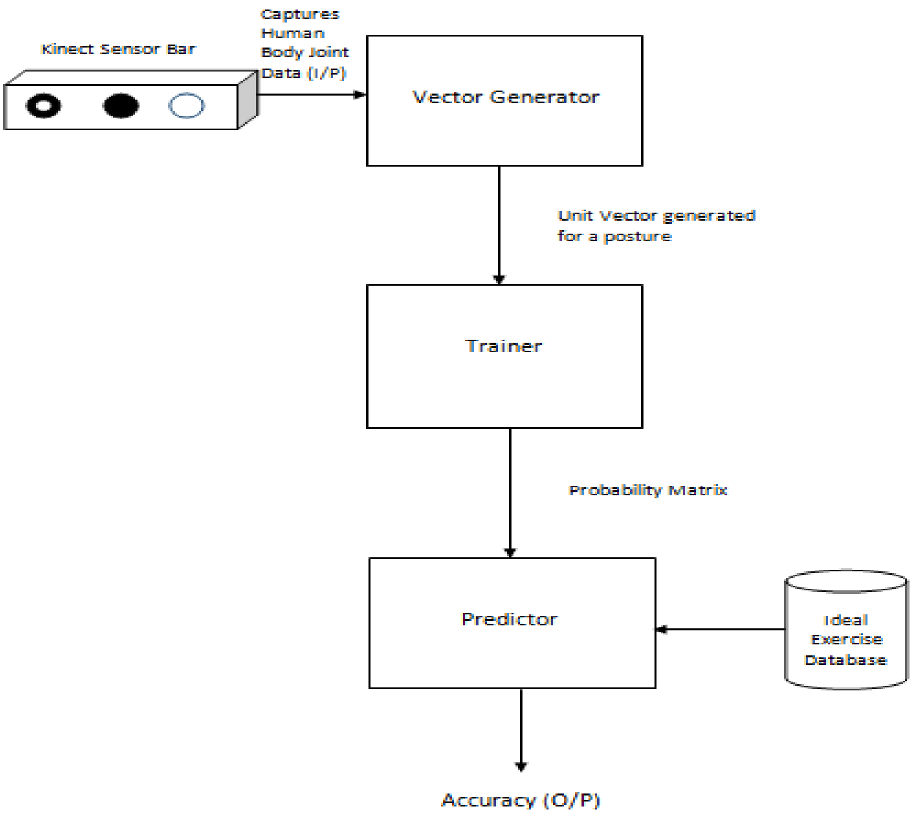


Figure 1: System Architecture

3.1 Kinect Sensor Bar

The Kinect sensor bar is an RGB-D camera introduces by Microsoft, which helps in developing gesture recognition systems by capturing depth information, and converting the captured information into 3D coordinates. The kinect sensor is shown in Figure 2.The kinect sensor consists of the following parts:-

1) **Infrared Sensor**
The IR sensor or the IR emitter is used to emit Infrared rays on to the subject who is performing the required gesture that is to be recognized by the system. The IR sensor helps capturing of the subject even in low light conditions. It works along with the depth camera.

2) **Depth Camera**
Along with the IR sensor, the depth camera helps capturing the depth images of the subject, for instance, the joint configuration data of a subject performing a particular gesture. Figure 3 shows a sample captured image by the depth camera.

3) **RGB Camera**
The RGB camera of the Kinect sensor is a normal camera, just like the other ones. The RGB camera has a 1280x960 resolution which is capable of storing three-channel data, thus making it possible to capture a color image

4) **Multi-Array Microphone**
A multi-array microphone contains four microphones for capturing sound. The four microphones make it possible to record audio as well as find the location of the sound source and the direction of the audio wave.

3.2 Vector Generator

The vector generator is responsible for the generation of unit vectors for the captured gestures. unit vector has a unit magnitude and has a main purpose to describe a direction in a 3D space. A unit vector in 3D space can be defined with three components which are along X, Y and Z axis, and can be derived as Let a vector V be unit vector from a point A to Point B, where (ax, ay, az) and (bx, by, bz) are the coordinates of points A and B respectively.

Then, component of vector V along X axis =

$$\frac{bx - ax}{\sqrt{(bx - ax)^2 + (by - ay)^2 + (bz - az)^2}}$$

Component of vector V along Y axis =

$$\frac{by - ay}{\sqrt{(bx - ax)^2 + (by - ay)^2 + (bz - az)^2}}$$

Component of vector V along Z axis =

$$\frac{bz - az}{\sqrt{(bx - ax)^2 + (by - ay)^2 + (bz - az)^2}}$$

The vector generator gives the output i.e., unit vectors for the gesture, as an input to the Trainer Module.

3.3 Trainer

The trainer is a module that is basically used for training the subject gesture data in order to generate the probability matrix for that gesture for the prediction of accuracy of the performed gesture with respect to the ideal gesture. The HMM machine learning technique is used for this module, where, at first, it received the unit vector of the gesture as an input. These unit vectors are used for the generation of the probability matrix of that gesture. The probability in the probability matrix is considered as the probability of a posture to be in that gesture. The detailed working of HMM is described in subsection 3.6. The output of the trainer module would be a probability matrix of a performed gesture, which is supplied as in input to the Predictor.

3.4 Ideal Exercise Database/Dataset

The ideal exercise database consists of a pre-generated ideal gesture database that is stored in the device on which the system works. The ideal exercises are also applied with HMM in order to generate a probability matrix for them. The determinants for each probability matrix are calculated, and these are stored in the database, which are later used by the predictor module for accuracy prediction.

3.5 Predictor

The predictor is the final module of the system, which is responsible for the final output of the system, which is the accuracy of the user performed exercise with respect to the ideal physical exercise in terms of percentage. Once the probability matrix for the user performed exercise is obtained, the determinant for that matrix is obtained and further used for accuracy prediction. Let the determinant for the user performed

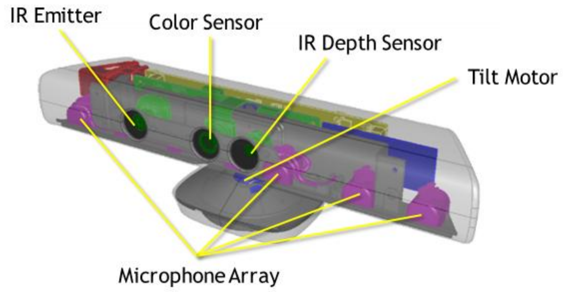
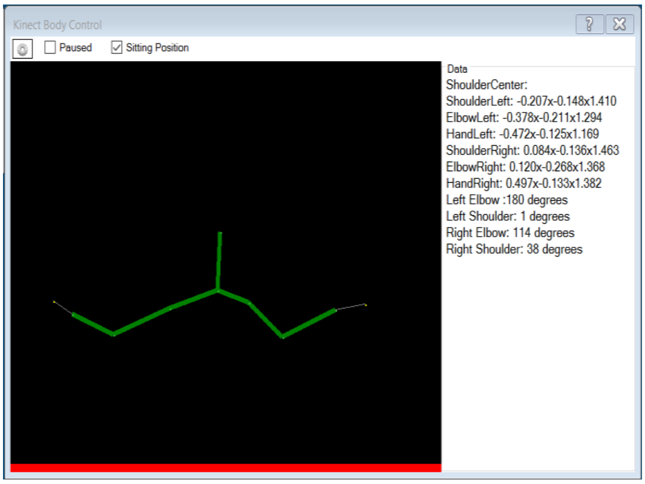


Figure 2: Kinect Sensor Bar



```
Data
ShoulderCenter:
ShoulderLeft: -0.207x-0.148x1.410
ElbowLeft: -0.378x-0.211x1.294
HandLeft: -0.472x-0.125x1.169
ShoulderRight: 0.084x-0.136x1.463
ElbowRight: 0.120x-0.268x1.368
HandRight: 0.497x-0.133x1.382
Left Elbow :180 degrees
Left Shoulder: 1 degrees
Right Elbow: 114 degrees
Right Shoulder: 38 degrees
```

Figure 3: Kinect Captured Data

The predictor is the final module of the system, which is responsible for the final output of the system, which is the accuracy of the user performed exercise with respect to the ideal physical exercise in terms of percentage. Once the probability matrix for the user performed exercise is obtained, the determinant for that matrix is obtained and further used for accuracy prediction. Let the determinant for the user performed

gesture be U and the determinant for ideal exercise be I. So the formulation for accuracy prediction in terms of percentage would be:-

$$\left[\frac{U - I}{I}\right] * 100\ \%$$

3.6 Algorithms Used

The system namely Virtual Personal trainer makes use of only one machine learning technique which is Hidden Markov Model(HMM). Hidden Markov Model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (i.e. hidden) states. This technique is mainly based on the concept of probability. The HMM takes the input as unit vector and is responsible for the generation of a probability matrix for the gesture, which can be used for recognizing the gesture, in this case, it would be further used for the prediction of accuracy of user performed gesture with respect to ideal gesture. The HMM works in the following way:-

Let the system have N distinct states. The different states within the system would be denoted as a set of states S :-

$$S_1, S_2 \ldots S_n$$

Let the state which would be considered at a particular time t as:-

$$q_t$$

So firstly, the Transition Probability Matrix A = a_{ij} which is the basic probability of a transition from state i to state j, which is given as:-

$$a_{ij} = P(q_t = S_j \mid q_{t-1} = S_i)\ ;\ i \geq 1 \& \ j \leq N$$

Next, the Observation Symbol Probability Matrix B = $b_j(k)$ which is basic probability of showing observation V_k in state j at time t:-

$$b_j(k) = P(V_k = S_j);\ 1 \leq j \leq N\ \&\ 1 \leq k \leq M$$

where M is the number of observations in the observation set.

3.7 Workflow

Figure 4 shows the overall working of the system via an activity diagram.

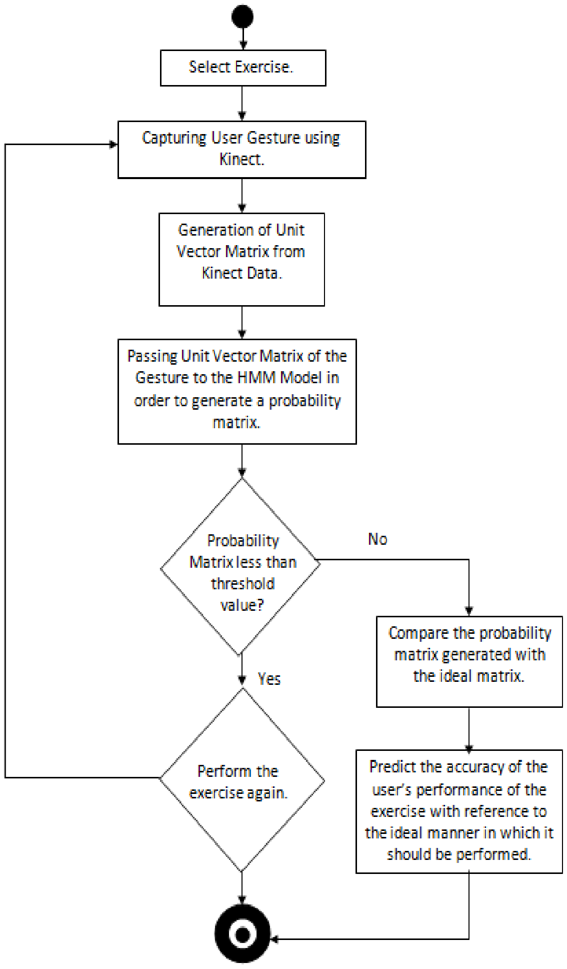

Select Exercise.
Capturing User Gesture using Kinect.
Generation of Unit Vector Matrix from Kinect Data.
Passing Unit Vector Matrix of the Gesture to the HMM Model in order to generate a probability matrix.
Probability Matrix less than threshold value?
No
Compare the probability matrix generated with the ideal matrix.
Yes
Perform the exercise again.
Predict the accuracy of the user's performance of the exercise with reference to the ideal manner in which it should be performed.

Figure 4: System Workflow via Activity Diagram

4. CONCLUSION

Thus, the paper proposes to develop a home training system which allows users to perform exercises under supervision anywhere, thereby, reducing the need for users to travel to the gym to perform basic routine exercises. This system promotes health and fitness amongst the youth and elderly alike. If used regularly, it serves as a guide and mentor to perform certain compound movements accurately, eliminating the possibility of injuries that may be caused due to incorrect form. The system is a very handy one with no pre-requisites of any kind, which allows beginners to use it with the utmost ease and convenience. There is a lot of future scope in terms of this system. While it definitely aims at increasing the motivation amongst people to work out and get fit, it can be developed and upgraded in such a way that it creates a personalized training model for individuals, suited to their lifestyle and needs. The Kinect Sensor can be used to identify bone and ligament injuries amongst sportsmen based on their movement during the game. Since proposed, this system is still under development and testing. Various tests were carried out on the sample data using various machine learning techniques like SVM, Random Forest Classifier, Dynamic Time Warping and HMM, of which HMM was found to give the maximum accuracy. Also the probability matrix generated as an output by the HMM technique is easy to be used for further computation. The system is expected to give accuracy almost similar to the accuracy of a personal trainer.

5. ACKNOWLEDGEMENTS

We would like to take this opportunity and express our sincere gratitude towards Prof. Rashmi.A.Rane for her guidance when required. We appreciate her valuable suggestions and support. We are also very grateful to Dr. V. Y. Kulkarni, Head of Computer Engineering Department, Maharashtra Institute of Technology, Pune for her tremendous support and guidance.

6. REFERENCES

[1] Salvatore Gaglio, Giuseppe Lo Re , Marco Morana Human Activity Recognition Process Using 3-D Posture Data. October 2015, IEEE Conference

[2] Lorenzo Patras, Ion Giosan, Sergiu Nedevschi Body gesture validation using multi-dimensional dynamic time warping on Kinect Data. September 2015, IEEE Conference

[3] Thomas Schlomer, Benjamin Poppinga, Niels Henze, Susanne Bol Gesture recognition using Wii Sensor. October 2017, IEEE Conference

[4] Geetha M, Manjusha C, Unnikrishnan P, Harikrishnan R A Vision Based Dynamic Gesture Recognition of Indian Sign Language on Kinect depth based images. October 2013, IEEE Conference

[5] Alina Delia Calin Gesture Recognition on Kinect Time Series Data using Dynamic Time Warping and Hidden Markov Model. September 2016, IEEE Conference

[6] Marcos Y.O Camada, Jes J.F Cerqueira, Antonio Narcus N. Lim Stereotyped Gesture Recognition:An Analysis between HMM and SVM. July 2017, IEEE Conference

[7] H. Haggag, M. Hossny, S. Nahavandi, O. Haggag An Adaptable System for RGB-D based Human Body Detection and Pose Estimation: Incorporating Attached Props. October 2016, IEEE Conference

[8] Kanad Biswas A Hidden Markov Model based Dynamic Hand Gesture Recognition System using OpenCV. 2015, Research Gate

[9] Rajat Srivastava Gesture Recognition using Microsoft Kinect. February 2013, IEEE Conference

[10] Hajar Hiyadi, Fakhreddine Ababsa, Christophe Montagne, El Houssine Bouyakhf, Fakhita Regragui Adaptive Dynamic Time Warping for Recognition of Natural Gestures . December 2016, IEEE Conference

[11] Muaaz Salagar, Pranav Kulkarni, Saurabh Gondane Implementation of Dynamic Time Warping for Gesture Recognition in Sign Language using High Performance Computing . August 2013, IEEE Conference

[12] Muhammad Hassan Khan, JullienHelsper, ZeydBoukhers, Marcin Grzegorzek Automatic Recognition of Movement Pattern in Vojta-Therapy using RGB-D Data. September 2016, IEEE Conference

[13] Sambit Bhattacharya, Bogdan Czejdo, Nicolas Perez Gesture Classification with Machine Learning using Kinect Sensor Data. December 2012, IEEE Conference

[14] Megha.D.Bengalur Human Activity Recognition Using Body Pose Features And Support Vector Machine. August 2013, IEEE Conference

[15] Zequn Zhang, Yuanning Liu, Ao Li, Minghui Wang A Novel Method for User-Defined Human Posture Recognition Using Kinect. October 2014, IEEE Conference

[16] Harshavardhan Verma, Eshan Agarwal , Satish Chandra Gesture Recognition Using Kinect for Sign Language Translation. December 2013, IEEE Conference

[17] Tie Yang, Yangsheng Xu Hidden Markov model For Gesture Recognition, May 1994.