IJCSN International Journal of Computer Science and Network, Volume 4, Issue 2, April 2015
ISSN (Online) : 2277-5420 www.IJCSN.org
**Impact Factor: 0.417**



# Real Time Hand Gesture Recognition System: A Review

[1] **Shaikh Shabnam,** [2] **Dr. Shah Aqueel**

[1] Department of Computer Engineering, Shri. Jagdish Prasad Jhabarmal Tibrewala University, Rajasthan, India

[2] Principal, Maulana Mukhtar Ahmad Nadvi Technical Campus, Malegaon

**Abstract -** The use of gesture as a natural interface serves as a motivating force for research in modelling, analysing and recognition of gestures. In particular, human computer intelligent interaction needs vision-based gesture recognition, which involves many interdisciplinary studies. A survey on recent real-time hand gesture recognition approaches is given in this paper. Hand gesture recognition is a relatively new field for the computer science. Applications for hand gesture recognition in machine learning systems have been developed approximately for 20 years. Research in motion-based recognition has greatly increased in last few years. Recently, there has been a surge of interest on hand detection, tracking, and gesture recognition. Human gesture recognition consists of identifying and interpreting automatically human gestures using a set of sensors (e.g. cameras, gloves). In this paper, efforts have been taken to present an up to-date review of the state-of-the-art in human gesture recognition which includes gesture recognition techniques, representations and applications.

**Keywords -** *Hand Gesture Recognition, Tracking, Fast Algorithm, Vision-based.*

## 1. Introduction

Recognizing gestures is a complex task which involves many aspects such as motion modelling, motion analysis, pattern recognition and machine learning, even psycholinguistic studies. There are already several survey papers in human motion analysis [1] and interpretation [2]. Since gesture recognition is receiving more and more attention in recent research, a comprehensive review on various gesture recognition techniques developed in recent years is needed. The idea was taken from Attila Licsar and Tamas Sziranyi [1]. This interface is simple enough to be run using an ordinary webcam and requires little training. The use of hand gestures provides an attractive alternative to cumbersome interface devices for human-computer interaction (HCI). In particular, visual interpretation of hand gestures can help in achieving the ease and naturalness desired for HCI. The main focus of the paper was on bare-hands utilizing a simple web camera with a frame rate of approximately 30 frames per second to communicate to computer all basic commands required by a human-computer interface. Digital video has become an integral part of everyday life. It is well-known that video enhancement is an active topic in computer vision has received much attention in recent years. The aim is to improve the visual appearance of the video, or to provide a “better” transform representation for future automated video processing, such as analysis, detection, segmentation, and recognition. Moreover, it helps analyze background Information that is essential to understand object behaviour without requiring expensive human visual inspection.

Video enhancement problem can be formulated as follows: give an input low quality video and the output high quality video for specific applications. The system can work with any camera that supports streaming video input to the computer. Its touch less interactive systems and mouse replacement solutions utilizes advanced computer vision to convert simple hand movements into direct mouse control in any environment. Asanterabi Malima et. al. [2] approach to the hand gesture recognition problem a robot control context involved the use of markers on the finger tips. An associated algorithm is used to detect the presence and color of the markers, through which one can identify which fingers are active in the gesture. The inconvenience of placing markers on the user’s hand makes this an infeasible approach in practice. Vafadar and Alireza Behrad[3] used Template Based: In this approach the data obtained is compared against some reference data and using the thresholds, the data is categorized into one of the gestures available in the reference data. This is a simple approach with little calibration but suffers from noise and doesn't work with overlapping gestures. Byung-Woo Min et. al. [4] developed Hidden Markov Model (HMM) which is commonly used and has been widely exploited for temporal gesture recognition. An HMM consists of states and state transitions with observation probabilities. For watch gesture a separate HMM is trained and the recognition of the gesture is based on the generation of maximum probability by a particular HMM.

This method also suffers from training time involved and complex working nature as the results are unpredicted because of the hidden nature. For the gesture recognition, Wing Kwong Chung et.al.[5] has presented a hand gesture recognition modal based on "A Real-time Hand Gesture Recognition based on Haar Wavelet Representation. In addition to voice and controller pads, hand gestures can also be an effective way of communication between humans and robots or even between auditory handicapped people and robots. Mu-Chun Su [6] suggested a method using Neural Network which is based on modelling of the human nervous system element called neuron and its interaction with the other neurons to transfer the information. Each node consists of and the input function which computes the weighted sum and the activation function to generate the response based on the weighted sum.

Byung-Woo Min et al.[7] developed a method for gesture recognition using Hidden Markov Model. This method has been widely exploited for temporal gesture recognition. Bhuyan, et.al. [8] have proposed the advantage of VOP based method for segmentation of hand image. The proposed acceleration feature works efficiently only when the spatial end position of preceding gesture is different from start position of next gesture in the connected gesture sequence. Shewta and Pankaj [9] have proposed that ANN provides a good and powerful solution for gesture recognition. Artificial Neural Networks are applicable to multivariate non-linear problems .It has fast computational ability. Gesture recognition is an important for developing alternative human-computer interaction modalities. Zhou Ren et.al. [10] have worked in the direction of hand gesture recognition by making use of kinetic sensor which is very much different from the normal web camera. Hand gesture based Human-Computer-Interaction (HCI) is one of the most natural and intuitive ways to communicate between people and machines, since it closely mimics how human interact with each other. Hamid A Jalab [11] have proposed the succeeds to extract features from hand gesture image based on hand segmentation using both wavelet network an ANN. Qing and Nicolas [12] proposed a method that used the formal grammar to represent the hand gestures and postures however limited. This method involves simple gestures requiring the fingers to be extended in various configurations which are mapped to the formal grammar specified by specific tokens and rules. The system involves tracker and glove. This system has poor accuracy and very limited gesture set. Lee and Yangsheng Xu [13] developed a glove-based gesture recognition system that was able to recognize 14 of the letters from the hand alphabet, learn new gestures and able to update the model of each gesture in the system in online mode, with a rate of 10Hz.

Over the years advanced glove devices have been designed such as the Sayre Glove, Dexterous Hand Master and Power Glove. Spatio-temporal vector Analysis method was proposed by Vafadar and Behrad [14] which used to track the movement of the hand in the images of the scene and track the motion in the sequence of image. The information about the motion is obtained by the derivatives and it is assumed that under static background, hand motion is the fastest changing object of the scene. Then using the refinement and variance constraint flow field is refined. This flow field captures the characteristics of the given gesture.

## 2. Classification of Gesture

Generally speaking, we can define a gesture as a body movement. A gesture is a non-vocal communication, used instead of or in combination with a verbal communication, intended to express meaning. It may be articulated with the hands, arms or body, and also can be a movement of the head, face and eyes, such as winking, nodding, or rolling eyes. Gestures constitute a major and an important mean of human communication. Figure 1 illustrates a taxonomy of gesture categories which resumes all the criteria except the last one.

This is not the unique way to classify gestures. Indeed, we can find in literature several other categorizations. However, we believe that the presented taxonomy accounts for almost all the aspects of gestures. In our work, we focus on dynamic gestures of the body and short actions independently from their semantic interpretation.

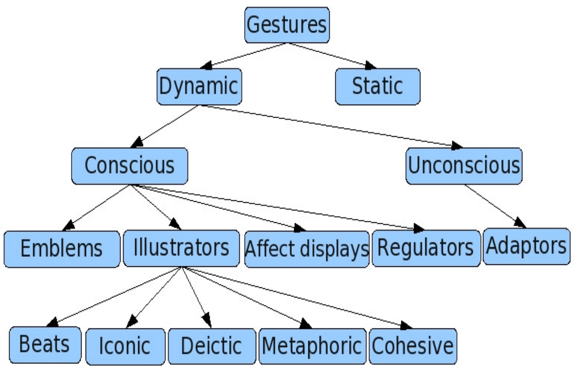


Fig. 1 Taxonomy of gesture categories

## 3. Gesture Recognition Techniques

There are two main kinds of devices for gesture recognition, Contact-based devices and Vision-based devices.

### 3.1 Contact-based Technology

Contact-based devices are various: accelerometers, multi-touch screen, instrumented gloves are, for instance, the main used technologies. Some devices, like Applec iPhon, include several detectors: multi-touch screen and an accelerometer for instance.

### 3.2 Vision-based Technology

Vision-based gesture recognition systems rely on one or several cameras in order to analyze and interpret the motion from the captured video sequences.

In a vision based hand gesture recognition system shown in Figure 2, the motion of the hand is captured by a camera . A set of features is extracted from every frame captured. A classifier uses the extracted features to recognize different postures for every frame. Since gestures are a dynamic sequence of hand postures connected though continuous movements, the classifier can be trained against a possible grammar. There are mainly two categories for vision based hand gesture recognition, namely the three dimensional (3D) hand model-based methods and, respectively, the appearance-based methods.

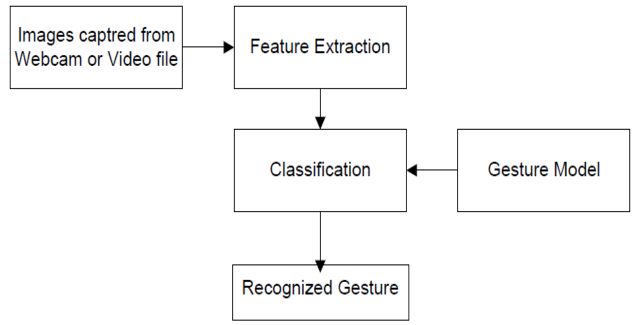


Fig.2 Vision-Based Hand Gesture Processing Stages

## 4. Gesture Representations

Several gesture representations and models have been proposed to abstract and model human body parts motion. We distinguish two main categories of method 3D model based methods and appearance based methods. Moreover, we can split the proposed models in two kinds according to the spatial and temporal aspects of gestures first is posture automaton models, in which the spatial and the temporal aspects are modelled separately and second one is motion models, in which there is a unique spatio-temporal model. Figure 3 overviews the different representations of gestures.

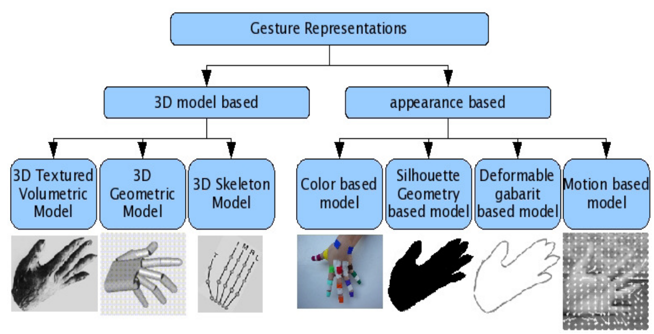


Fig. 3 Different representations of gesture

### 4.1 3D-Model Based Methods

A 3D model defines the 3D spatial description of the human body parts. The temporal aspect is generally handled by an automaton which generally divides the gesture time into 3 phases (1) the preparation or pre-stroke phase, (2) the nucleus or stroke phase and (3) the retraction or post-stroke phase. Each phase can be represented as one or several transition(s) between the spatial states of the 3D human model. The main advantage of 3D model based methods is to recognize gestures by synthesis: during the recognition process, one or more camera(s) are looking at the real target and then compute the parameters of the model that matches spatially the real target and then follows the latter motion (i.e. update the model parameters and check whether it matches a transition in the temporal model). Thus, the gesture recognition is generally precise (specially the start and the end time of the gesture). However, these methods tend to be computationally expensive unless implemented directly in dedicated hardware. Some methods combine silouhette extraction with 3D model projection fitting by finding the target self-orientation. Generally, three kinds of model are usually used:

*Textured kinematic/volumetric model:* these models contain very high details of the human body: skeleton and skin surface information.

*3D geometric model:* these models are less precise than the formers in terms of skin information but still contain essentially skeleton information.

*3D skeleton model:* these are the most common 3D models due to their simplicity and higher adaptability: The skeleton contains only the information about the articulations and their 3D degree of freedom (DOF).

### 4.2 Appearance-based Methods

Concerning appearance-based methods, two main sub-categories exist:

(1) 2D static model based methods and
(2) motion-based methods. Each sub-category contains several variants. For instance, the most used 2D models are:

*Colour-based models***:** methods with this kind of model use generally body markers to track the motion of the body or the body part. For example, a method for hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering.

*Silhouette geometry based models:* such models may include several geometric properties of the silhouette such as perimeter, convexity, surface, compactly, bounding box/ellipse, elongation, rectangularity, centroid and orientation. The geometric properties of the bounding box of the hand skin to recognize hand gestures.

*Deformable gabarit based models:* they are generally based on deformable active contours They used snakes for the analysis of gestures and actions in technical talks for video indexing. For motion models, we can split them in two variants:

Global motion descriptor: An "action sketch" extracted from a silhouette motion volume obtained by stacking a sequence of tracked 2D silhouettes. The "action sketch" is composed of a collection of differential geometric properties (e.g. peak surface, pit surface, ridge surface) of the silhouette motion volume. For recognizing an action, the authors use a learning approach based on a distance and epi-polar geometrical transformation for viewpoint changes.

## 5. Approaches for Human Gesture Recognition

Early approaches in gesture recognition from video sequences focused mainly on optical flow and motion history analysis. Throughout the past two decades, the number of techniques has been increased and different categories of approaches emerged. In this section, we outline the different methods of human gesture recognition from video sequences. Figure 4 shows some common approaches for human gesture identification.

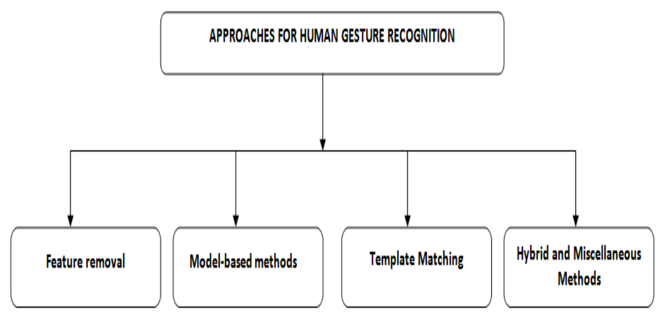


Fig.4 Approaches for human gesture recognition

### 5.1 Feature Extraction and Statistical Classification

These methods are generally associated with a global or local motion model. The features (i.e. motion descriptors) are extracted from the input video sequence. Two stages are necessary for the recognition: (1) a learning stage where the extracted features from a training video dataset are categorized and (2) a classification stage where the extracted features are compared to the learned ones.

### 5.2 Model-based Methods

As seen in previous section there are two types of gesture model: (1) 3D models and (2) 2D models. Unlike learning based methods, the recognition process is composed of a single stage where the parameters of the real target are extracted and then fitted to the adequate gesture model. The model fitting is driven by the attempt to minimize a residual measure between the projected model and the person contours (e.g. edges of the body) which requires a very good segmentation of body parts. Thus, such techniques require video sequences without very strong noise. A method for gesture recognition using 3D Visual Hull (3D geometric model) approach reconstructs the 3D human body model from several point of views captured by several cameras (for instance four cameras). Then a 3D shape descriptor consisting of a set of geometric properties of the 3D model is computed. Finally, "matching pursuit decomposition" is performed using a priori known 3D model of postures and a dual-state HMM.

### 5.3 Template Matching

In template matching based methods there is neither a feature extraction nor a gesture model, the whole gesture is considered as a template. The motion history images (MHI) and motion energy images (which can be seen as gesture templates) to recognize gestures. During the learning process, a statistical model (mean and covariance matrix) of seven Hu moments of MHI and MEI is generated. For the matching process, a Mahalanobis distance is computed between the moment description of the new input and moments of learned gestures. Recently, a volume motion template (VMT) for view-invariant gesture recognition. The training data-set consists of the projection images of the VMTs of each learned gesture. For matching a new input gesture, the k-nearest neighbour algorithm is used. These methods are different from 2D methods since the temporal and spatial dimensions are included in the same model. So, there is no need for an automata-like recognizer. Also, this is different from motion-based models where local or global motion descriptors (e.g. HOG, SIFT) are extracted and then learned. Here, the whole gesture template (or a more compact representation) is used as a learning model. The disadvantage of such methods is the huge size of learned data which influences the computational cost of the matching process.

### 5.4 Hybrid and Miscellaneous Methods

A combine a depth silhouette (3D model) to a temporal template matching method define a multi-layered silhouette history image (MLSHI). Each layer corresponds to motion history of one silhouette depth. Each template layer is reduced to its principal components using PCA. Then, the component is concatenated to feed a SVM classifier. Author introduces an action recognition method using Finite State Machines (FSM) and a Hybrid posture recognition process. A 3D posture avatar which models several postures of the human body. In the recognition process, the person orientation is estimated and then the 2D silhouette is compared to the projections of the a priori known postures from the 3D avatar according to the computed orientation. The silhouette comparison is done using different appearance based techniques (e.g. horizontal and vertical projections). The action recognition is performed by using a FSM in which each state represents one or several postures.

## 6. Application Areas of Gesture Recognition

**G**esture recognition has several domains of application. From sign language interpretation to virtual environments with smart human-machine interfaces, the number of domains increases continuously and the proposed solutions are more and more efficient.

### 6.1 Sign Language

Sign languages are the most raw and natural form of languages could be dated back to as early as the advent of the human civilization, when the first theories of sign languages appeared in history. It has started even before the emergence of spoken languages. Since then the sign language has evolved and been adopted as an integral part of our day to day communication process. Now, sign languages are being used extensively in international sign use of deaf and dumb, in the world of sports, for religious practices and also at work places

### 6.2 Presentations or Gesture-to-Speech

In a presentation (e.g. weather narration, technical reports), significative gestures are recognized to identify more precisely what the speaker is talking Specially, we can be interested to detect the directional indication through pointing. This is useful for identifying the context of statements or instructions. Moreover, it can be very interesting to interpret gestures at the same time with the speaker speech since there is some correlation between them.

#### 6.2.1 Virtual Reality

Virtual reality (i.e. virtual environments) allows a user to interact with a computer-simulated environment (whether that simulated world is an instance of the real world or an instance of imaginary world). It includes immersive gaming, flight simulators and remote control. Here gesture recognition is used as a mean of communication with the virtual world (eventually the real system).Virtual reality is applied to computer-simulated environments that can simulate physical presence in places in the real world, as well as in imaginary worlds.

### 6.2.2 Medical Systems and Assistive Technologies

Hand gestures can be utilized to interact with visualization screens, control medical devices, and assist handicapped persons as part of their rehabilitation treatment. Some of these ideas have been used to enhance medical procedures and systems. In some methods it is discussed to integrate hand gestures with physician-computer interfaces, illustrating a computer-vision system that allows doctors to carry out standard mouse functions such as pointer movement and button presses with hand gestures that meet with the "intuitiveness" requirement.

## 6.3 Video Surveillance

The main objectives of video surveillance are security and safety. The first thoughts that spring immediately to mind are about the security issue. As for gesture recognition, The main goal is to detect violent and/or furtive actions in some areas to secure. The faces of violent actors are then recognized. Concerning the second objective (i.e. safety) of video surveillance, the goal is to check if the gestures and the actions of a certain operator are safe in a certain zone.

## 6.4 Gaming (Entertainment)

Computer games are a technologically promising and commercially rewarding field for innovative interfaces because of the entertaining nature of the interaction. People are willing to try new interface technologies because they have the chance to be immersed in a challenging game-like environment. In computer-vision-based hand gesture- controlled games, the system has to react fast to user gestures, must be robust and efficient in contrast to other applications such as inspection systems, which do not need real-time requirement.

## 6.5 3D Design

CAD (computer aided design) is an HCI which provides a platform for interpretation and manipulation of 3-Dimensional inputs which can be the gestures. Manipulating 3D inputs with a mouse is a time consuming task as the task involves a complicated process of decomposing a six degree freedom task into at least three sequential two degree tasks. Massachuchetttes institute of technology has come up with the 3DRAW technology that uses a pen embedded in polhemus device to track the pen position and orientation in 3D.A 3space sensor is embedded in a flat palette, representing the plane in which the objects rest .The CAD model is moved synchronously with the users gesture movements and objects can thus be rotated and translated in order to view them from all sides as they are being created and altered.

## 6.6 Human-Robot Interaction

Another traditional topic of application for gesture interaction systems is human-robot interaction with a huge number of published papers. Hand gesture recognition is a significant feature for fixed and portable robots, robotic manipulators and human users were integrated with hand gesture commands for recognizing four postures such as when a user points at an object, the robot can detect it. A human-machine interaction system for game playing allows the visual recognition of three postures with different rotation angles and scales with 95% accuracy. A robotic- assistant interaction application was designed by combining gesture recognition and voice to generate commands. Once the hand posture is detected, six gestures are trained using a hand contour as the major feature of each gesture. A human-robot system was designed to recognize dynamic gestures by tracking head and arm direction. The system employs Hidden Markov Model (HMM) to recognize trajectories of the detected hands. A programming-by-demonstration method allows the robot to learn a human user's gestures. The system employs eight hand postures to control a hybrid service robot system. A programming-by-demonstration idea was also utilized to allow users to help the robot make gestures by kinaesthetic teaching. The user trains the robot to make 10 dynamic gestures captured by sensors connected to the user's torso, upper- and lower-arm. There are so many amount of studies found in literature survey for the hand gesture recognition system. Recognition methods, like the detection procedure, that mainly rely on algorithms that need training or having diverse environmental constraints. The summary of such algorithms are shown in table 1.

Table 1: Gesture Recognition Methods

| **Primary Method of Recognition** | **No. of Gestures Recognized** | **Background to Gesture Images** | **Additional Markers Required** | **Number of Training Images** |
|---|---|---|---|---|
| Hidden Markov Models | 97 | General | Multicolored gloves | 400 |

| | | | | |
|---|---|---|---|---|
| Entropy Analysis | 6 | No | No | 400 |
| Linear approximation to non-linear point distribution models | 26 | Blue Screen | No | 7441 |
| Finite State machine modeling | 7 | Static | Markers on glove | 10 sequences of 200 frames each |
| Fast Template Matching | 46 | Static Wrist band | 100 | examples per gesture |

## 7. Conclusion

In summary, a review of vision-based hand gesture recognition methods has been presented. Considering the relative infancy of research related to vision-based gesture recognition, remarkable progress has been made. To continue this momentum, it is clear that further research in the areas of feature extraction, classification methods and gesture representation are required, to realize the ultimate goal of humans interfacing with machines on their own natural terms. In this paper the recent development on the research of hand gesture recognition with focus on various recognition techniques. Overall, gesture recognition is still in its infancy. It involves the cooperation of many disciplines. In order to understand hand gestures, not only for machines, but also for humans, substantial research efforts in computer vision, machine learning and psycholinguistics will be needed.

## References

[1] Attila Licsar and Tamas Sziranyi, IEEE Int, “Hand Gesture based film Restoration ACM Cybernetics” pp. 215-22, 2004.

[2] Asanterabi Malima, Erol Özgür, and Müjdat Çetin “A Fast Algorithm For Vision-Based Hand Gesture Recognition for Robot Control”IEEE Int, Conf. Systems, Man, Cybernetics. pp. 289-296, 2008.

[3] Vafadar and Alireza Behrad “Human Hand Gesture Recognition Using spatio- Temporal Volume for Human Computer Interaction” IEE Int. Symposium on Telecommunications, pp.713-718,2008.

[4] Byung-Woo Min, Ho-Sub Yoon, Jung Soh, Yun-Mo Yang and Toshiaki Ejima, “Hand Gesture Recognition Using Hidden Markov model” IEEE Int Conf. systems,Man, and Cybernetics, 1997 , pp 4232-4235.1997.

[5] Wing Kwong Chung, Xinyu Wu and Yangsheng Xu, “A Realtime Hand Gesture Recognition based on Haar Wavelet Representation” Proceedings of the 2008 IEEE Int. Conf. on Robotics and Biomimetics , pp.336-341,2008.

[6] Mu-Chun Su “A Fuzzy Rule-Based Approach to Spatio-Temporal Gesture Recognition” IEEETrans on systems, Man, And Cybernetics –Part C: Applications and Reviewws, Vol.30No 2, pp276-281,2000.

[7] Byung-Woo Min, Ho-Sub Yoon, Jung Soh, Yun-Mo Yang and Toshiaki Ejima, “Hand Gesture Recognition Using Hidden Markov model” IEEE Int. Conf. systems,Man, And Cybernetics, 1997 , pp 4232-4235,1997.

[8] M K Bhuyan, D Ghosh, P K Bora,“Feature Extraction from 2D Gesture Trajectory in Dynamic hand Gesture Recognition” 1–4244–0023–6/06/ IEEE pp. 432-441,2006.

[9] Shewta Yewale ,Pankaj Bharne,“Artificial neural network approach for hand gesture recognition” International journal of engineering Science and technology, Vol 3 No.4,pp. 2603-2608,2011.

[10] Zhou Ren Jing jing Meng Junsong Yuan,Zhengou Zhang, "Robust hand gesture recognition with kinect Sensor” ACM multimedia 2011 conference ,pp. 978-979.2011.

[11] Hamid A Jalab,“Static hand Gesture recognition for human computer interaction” Asian Network for Scientific Information information technology journal, pp. 1-7,2012.

[12] Qing Chen, Nicolas D. Georganas, “Hand Gesture Recognition Using Haar-Like Features and a Stochastic Context-Free Grammar”, IEEE transactions on instrumentation and measurement, vol. 57, no. 8, pp.1562-1571,2008.

[13] Christopher Lee and Yangsheng Xu, “Online, interactive learning of gestures for human robot interfaces” Carnegie Mellon University, The Robotics Institute, Pittsburgh, Pennsylvania, USA,1996.

[14] Vafadar,Alireza Behrad,“Human Hand Gesture Recognition Using spatio-Temporal Volume for Human Computer Interaction” IEE Int. Symposium on Telecommunications, pp.713-718,2008.