JOURNAL OF COMPUTING TECHNOLOGIES

ISSN 2278 – 3814

Available online at www.jctjournals.com

Volume 1 , Issue 2 , June 2012

Bare Hand Gesture Recognition-A study

Kashmera Khedkkar Safaya[#1], Asst.Prof.Rekha Lathi[#2]

[#]*Information Technology, Mumbai University*
Sector-9 RH-2 S-13 C.B.D Navi Mumbai- 400416, India
[1]kashmera.k@gmail.com

[*]*Computer Department, Mumbai University*
[2]rekha_lathi@yahoo.com

***Abstract*— This paper proposes a method to recognize bare hand gestures using a dynamic vision sensor (DVS) camera. DVS cameras only respond to pixels with temporal luminance differences, which can greatly reduce the computational cost of comparing consecutive frames to track moving objects. This paper attempts to classify three different hand gestures. We propose novel methods to detect the delivery point, to extract hand regions, and to extract useful features for machine learning based classification. In order to do this, the paper begins by trying to understand the importance of gestures and how humans use gestures to communicate.**

***Keywords*— Dynamic vision sensor camera, Hand gesture recognition**

I. HUMANS AND GESTURE

Humans use various forms of expression like speech, facial expression, and bodily movements to communicate. Humans and gestures go back a long way. Gestures are ingrained so deeply in humans that they are considered part of language and expression. An example of this is clearly seen when people use hand gestures when they talk on the phone even though they know that they cannot be seen. Due to the widespread use, and importance of gestures in everyday human interaction, a natural extension would be to incorporate some aspects of gestural interaction in computer systems. It is hard to settle on a specific useful definition of gestures due to its wide variety of applications and a statement can only specify a particular domain of gestures. Many researchers had tried to define gestures but their actual meaning is still arbitrary. Bobick and Wilson have defined gestures as the motion of the body that is intended to communicate with other agents. For a successful communication, a sender and a receiver must have the same set of information for a particular gesture. As per the context of the project, gesture is defined as an expressive movement of body parts which has a particular message, to be communicated precisely between a sender and a receiver. A gesture is scientifically categorized into two distinctive categories: dynamic and static. A dynamic gesture is intended to change over a period of time whereas a static gesture is not intended to change over the time. A waving hand means good bye is an example of dynamic gesture and the stop sign is an example of static gesture. To understand a full message, it is necessary to interpret all the static and dynamic gestures over a period of time. This complex process is called gesture recognition. Gesture recognition is the process of recognizing and interpreting a stream continuous sequential gesture from the given set of input data.

II. MOTIVATION

A review of existing literature shows that there has been a considerable amount of work done in the area of gesture based interactions, and its application in various domains. In addition, quite a bit of research exists around how gestural inputs are recognized and processed, and on the results of such gestural interactions. However, there seems relatively little in the way of best practices for gestures, and the qualities

that make for a good gesture. This paper makes an attempt to fill this void.

III. LITERATURE REVIEW

A number of researchers have explored the use of hand gestures as a means of computer input, using a variety of technologies. Research has been limited to small scale systems able to recognize a minimal subset of a full sign language. Christopher Lee and Yangsheng Xu 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. Over the years advanced glove devices have been designed such as the Sayre Glove, Dexterous Hand Master and PowerGlove. The most successful commercially available glove is by far the VPL DataGlove as shown in fig 1.

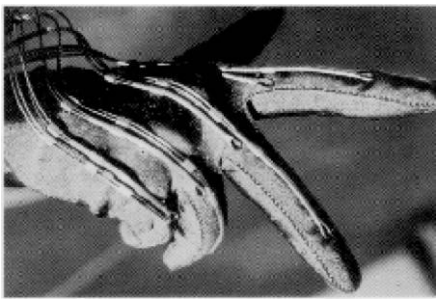


Fig.1 VPL data glove

It was developed during the 1970″s. It is based upon patented optical fiber sensors along the back of the fingers. In later years a glove-environment system is designed and it was capable of recognizing 40 signs from the American Sign Language (ASL). Hyeon-Kyu Lee and Jin H. Kim presented work on real-time hand-gesture recognition. Kjeldsen and Kenders devised a technique for doing skin-tone segmentation in color model.. They further developed a system which used a back-propagation neural network to recognize gestures from the segmented hand images. Etsuko Ueda and Yoshio Matsumoto presented a novel technique a hand-pose estimation that can be used for vision-based human interfaces, in this method, the hand regions are extracted from multiple images obtained by a multiviewpoint camera system, and constructing the "voxel Model." Hand pose is estimated. Chan Wah Ng, Surendra Ranganath presented a hand gesture recognition system, they used image furrier descriptor as their prime feature. Their system″s overall performance was 90.9%. Claudia Nölker and Helge Ritter presented a hand gesture recognition modal based on recognition of finger tips, in their approach they find full identification of all finger joint angles. Implementation issues in gesture systems include intention/immersion, recognition and segmentation of gestures. Intention is the problem of deciding whether the user is genuinely gesturing or not; avoiding false positives from recognizing casual motions as gestures. The term immersion is used to describe the problem of the user″s hand being always under analysis. This is a feature of gesture recognition systems that is not experienced in conventional mouse and keyboard interaction, where the user can simply let go, completely disengaging themselves from interaction. Recognition is identifying the user″s intended gestures, while segmentation is determining where one gesture ends and another begins in a command sequence. In 1989 Sturman et al used a data glove, capable of measuring hand position, hand orientation and flex angles for the thumb and each finger, to allow manipulation of objects in a virtual world. They considered three ways in which gesture input might be used. The first is to directly manipulate objects by „reaching into″ an environment and holding and moving. They, however, found out that the absence of tactile feedback caused some difficulties. The second use of gestures by Sturman et al was to operate abstract input devices: buttons, valuators and locators, positioned in a three dimensional volume. In their system hand static postures and motion onset were used as buttons to trigger actions. Experiments with finger flexing found that people did not readily issue static gestures depending on accurate finger angles.

IV THE GESTURE RECOGNITION SYSTEM

Different methods and technologies have been used for gesture recognition, but the major steps in gesture recognition are the same. The complete system is categorized in to three parts,

A. Hand Segmentation

This step is also known as hand detection. It involves detecting and extracting hand region from background and segmentation of hand image. Different features such as skin colour, shape, motion and anatomical models of hand are used in different methods. The output of this step is a binary image in which skin pixels have value 1 and non-skin pixels have value 0. Different methods for hand detection are summarized in this paper. Some of them are.

Colour: Different colour models can be used for hand detection such as YCbCr, RGB, and YUV etc.

Shape: The characteristics of hand shape such as topological features could be used for hand detection. Learning detectors from pixel values: Hands can be found from their appearance and structure such as Adaboost algorithm. 3D model based detection. Using multiple 3D hand models multiple hand postures can be estimated.

B. Feature Extraction

The next important step is hand tracking and feature extraction. Tracking means finding frame to frame correspondence of the segmented hand image to understand the hand movement. Following are some of the techniques for hand tracking.

a) Template based tracking: If images are acquired frequently enough hand can be tracked. It uses correlation based template matching. by comparing and correlating hand in different pictures it could be tracked.

b) Optimal estimation technique: Hands are tracked from multiple cameras to obtain a 3D hand image.

c) Tracking based on mean shift algorithm: To characterize the object of interest it uses colour distribution and spatial gradient. Mean shift algorithm is used to track skin colour area of human hand. Two types of features are there first one is global statistical features such as centre of gravity and second one is contour based feature that is local feature that includes fingertips and finger-roots. Both of these features are used to increase the robustness of the system.

C. Gesture Recognition

The goal of hand gesture recognition is interpretation of the meaning of the hands location and posture conveys. From the extracted features multiple hand gestures are recognized. Different methods for hand gesture recognition can be used such as template matching, method based on principle component analysis, Boosting contour and silhouette matching, model based recognition methods, Hidden Markov Model (HMM). Hand gesture is movement of hands and arms used to express an idea or to convey some message or to instruct for an action. From psychological point of view hand gesture has three phases.

a) Preparation: bringing hand to starting posture of gesture.

b) Nucleus: includes main gesture.

c) Retraction: this includes bringing hand to resting position.

Finding starting and ending position of the nucleus phase is a difficult task because different persons have different shape and hand movement.

D. Hand Modelling

Different model based solutions for hand gestures are proposed here. A typical vision based hand gesture recognition system consists of a camera, a feature extraction module, a gesture classification module and a set of gesture models. The necessary hand features are extracted from the frames captured by video. These features are classified as,

a) High level features that are based on three dimensional models.

b) The image is used by view based approach as a feature.

c) Some low level features that are measured from image.

Hand gesture is the movement of hand in air so it is required to define a spatial model to represent these movements. Two types of models are used. Volumetric models that describes the shape and appearance of hand and skeletal model represents hand posture.

V. CLASSIFICATION OF VISION BASED GESTURE RECOGNITION METHODS

There are a number of methods that are used for hand gesture recognition. Some of the vision based hand gesture recognition systems are discussed below.

A. Hand Modelling with High Level Features:

In this method multiple images are captured by multiview point camera and then the hand regions are extracted from images. Using all these multiview point images and integrating them a hand posture can be constructed. this model is compared with hand model to recognize hand posture with the help of hand tracking.

B. View Based Approach:

The hand is modelled using a collection of 2D intensity images and the gestures are modelled as a sequence of views. These approaches are also called appearance-based approaches. This approach has been successfully used for face recognition. Eigenspace approaches are used within the view-based approaches. They provide an efficient representation of a large set of high dimensional points using a small set of orthogonal basis vectors. These basis vectors span a subspace of the training set called the eigenspace and a linear combination of these images can be used to approximately reconstruct any of the training images.

C. Low Level Features:

This method is based on the assumption that detailed information about hand shape is not necessary for humans to interpret sign language. It is based on the principle that all humans hand has approximately same hue and saturation. A low level feature set is used to recognize hand posture. This method achieves approximately 97% accuracy.

D. Gesture Segmentation Method Based On Complexion Model:

In this method a gesture segmentation method based on complexion model and background model that uses Fourier Descriptor and BP neural network is discussed. Hand segmentation is done by selecting colour space and building complexion model and background model. The only use of complexion model may produce some interference by similar complexion region. This interference can be removed by using background model along with complexion model. In this method the gesture is recognized using Fourier descriptor and BP neural network. In this approach contours are described effectively by using Fourier descriptor because it is rotational invariance, translation invariance and scale invariance. Y calculating fourier factors of the border points of the gesture fourier descriptors can be obtained. Hand gesture can be identified fast with this method. Hand gesture classification based on BP neural network solves the problem of low recognition rate and problem of gesture segmentation in intricate background. The experimental results show that the method is flexible, realistic, exact and fit for many applications in virtual reality. But in high light and shadow the result is still not perfect.

Fig. 2(a)

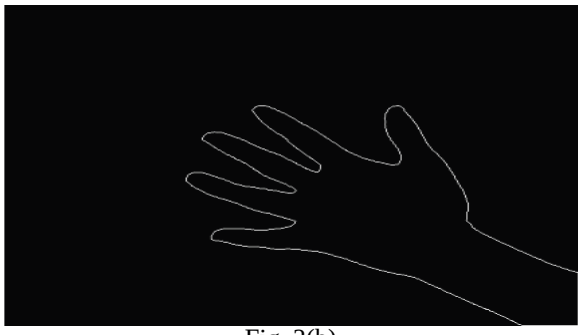


Fig. 2(b)

Fig.2(c)

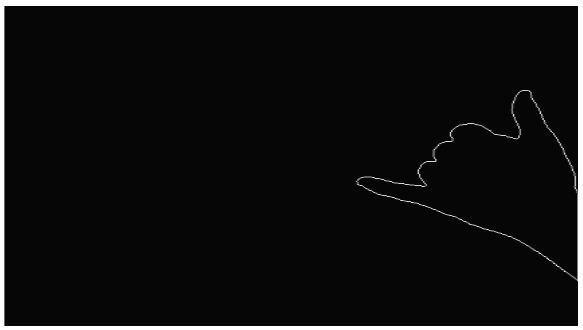


Fig. 2(d)

Fig. 2 Segmentation result. (a) and (c) original picture and fig (b) and (d) segmentation result [7]

E. Multihand Posture Recogition Based on Haar like Feature and Topological Features:

This method uses haar like feature detector and topological method along with color segmentation to get accurate hand posture recognition. A rich set of Haar-like features are computed from the integral image. Each Haar-like feature is composed of two connected "black" and "white" rectangles. The value of a Haar-like feature is obtained by subtracting the sum pixel values of the white rectangle from the black rectangle. Hand region segmentation: this is an important step in hand posture recognition. This is accomplished by two important techniques, haar-like feature and color based model. Haar like detector is used to detect both left and right hand posture. Initially the input image is transformed in to an integral image because from integral image these features could be extracted fast. A set of haar like features are computed from the integral image. Edge and rotated haar like features were proposed in this algorithm that gives better description of hand posture. Therefore more than 60,000 features could be extracted from each sub of input image with size of 24×30

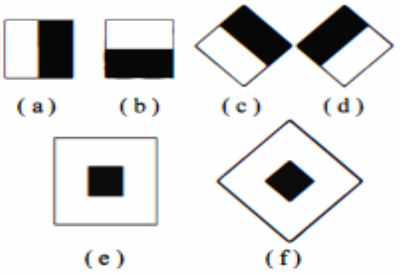


Fig. 3 Haar-like features[7]

Figure 3: Haar-like features

For feature extraction following steps are followed

a) Hand region area is converted to binary image. Hand area has pixel value 1 and nonhand region has value 0

b) Centre point of hand palm is achived Co by eroding the hand. Calculate maximum distance from centre of hand to hand region boundary Dn. Taking Co as the centre point a search circle is drawn with radius Ri=Dm/10.

c) From the pixel value of point along the circle and record the value of all change points.

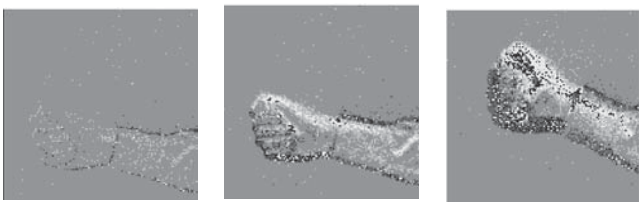
d) Repeat steps 2 and 3 by increasing redius Ri+1=Ri+Dm/10 until Ri=Dm.

VI. PROPOSED SYSTEM

The proposed system makes use of best of these techniques this system have following steps

A. Detecting Delivery Point

The delivery point refers to a point where it delivers a throw. The movement of the hand is slower as the hand reaches the end of the delivery point in order to stop moving at the delivering phase. DVS cameras are only response to the pixels with luminance intensity, and send events of only those pixels. Since there is little movement of the hand at the delivery point, the number of events within a frame will be dramatically reduced. Once the delivery point is detected, system can classify different throws using a single posture captured at the time of delivery point. To find the delivery point, track the number of events for each frame, and if the number of events in the frame is less than the given threshold TH, the frame is regarded as a delivery point. If TH is too large, the frame detected as a delivery point can be located at the middle of the approach phase. In this case, the hand pose in the detected frame can be different from the actual hand throw. Even if the hand pose in the detected frame is the same as the actual throw, the frame may not be the appropriate one for detecting hand pose. The reasons are that the frame is likely to contain too many events, which will increase the computational cost, and the shape of the hand can be not clear by thickening the boundary as shown in fig. 4(c).In contrast, if the size of TH is too small, as in fig. 4(a) the frame detected as a delivery point may contain too few events to recognize the shape of hand. Therefore, TH needs to be carefully chosen. Fig.4 (a) (b) and (c) show the frames detected as a delivery point.



a) TH =1,000 b) TH=5,000 c) TH=10,000

Fig.4 Frames at the delivery points

However, stopping the hand movement does not necessarily mean that the player is delivering a throw. Therefore it track the direction of movement for each frame, and the first frame with number of events less than TH during the downward movement is considered as delivery point. To summarize, the frame is regarded as a delivery point when it has the number of events less than the given threshold (TH) and the hand is moving downward, where the moving direction of the hand is determined as delivery point.

B. Hand Posture Recognition

Once the delivery point is detected, the frame at the delivery point is used to recognize hand pose. Since there are some noisy events, a noise filtering process is first conducted using connected component analysis. It is often useful to extract regions which are not separated by a boundary. [3] Any set of pixels which is not separated by a boundary is call connected.To apply connected component analysis, the stream of events of a frame is represented by a 128 by 128 matrix. There are two types of events: on- and off- events. Although a single matrix can be generated by ignoring the event types each of which represents on- and off- events are used, and connected component analysis is applied to each matrix separately. Since events of the same type are likely to occur closely in space, we expect that the two matrix representation can filter out noisy events that are surrounded by events of a different type.

C. Hand Extraction

After the filtering process, the hand region is first extracted. Some portion of the forearm is presented in the frame. Since the various lengths of forearm presented in the frame do not provide useful information to distinguish different hand postures. To do this, first estimate the width of the hand along the horizontal axis, and then record changes in width from right to left (from forearm area to hand) to locate the point with the largest width increase. Hand postures, only the hand region is used for classification. To extract the hand, first the wrist point is found. It works as follows:

The frame is segmented into b_H bins along the horizontal axis as shown in fig. 5, and the segmentation point which lies in between the *i*th bin and the *i*+1 bin is denoted as ∂_i The size of the bin is $|max(x) - min(x)| / b_H$, where *max(x)* and *min(x)* represent the largest and the smallest x-address of events respectively, and b_H is the number of bins. Estimate the width of the hand for each bin, which is denoted by d_i, for *i*th bin. Calculate the width change for each segmentation point ∂_i by subtracting the width of the hand in the left bin of ∂_i i.e. d_i-d_{i+1}. From forearm area to hand, find the segmentation point *p* with the maximum changes. All the events whose x-address is smaller than *p* are regarded as a hand region. In addition, the number of bins, b_H, should be appropriately set. If b_H is too small, the actual width changes of the hand cannot be captured during the hand extraction process, and the algorithm is likely to detect a point which is not even close to the actual wrist as a wrist point. Even if the detected wrist point is near the actual wrist, the extracted hand region can contain a large portion of forearm or exclude a large part of the actual hand region due to a large bin size.

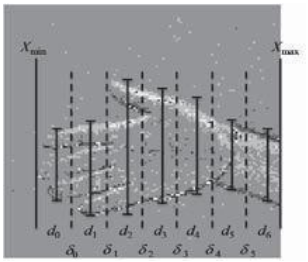


Fig.5. Hand extraction [1]

D. Feature Extraction

Feature extraction is very important in terms of giving input to a classifier. The aim of this phase is to find and extract features that can be used to determine the meaning of a given gesture. One simple feature which can represent the shape of a hand might be the width distribution across the hand. Similar to the method to extract the wrist point, we segment the extracted hand into b_F bins along the horizontal axis as shown in the fig. 6, and then estimate the hand width for each bin. A sequence of hand widths shows how the size of width changes along the horizontal axis. Since the absolute value of width can be different depending on person and the distance between the hand and the camera we use relative width, which can be obtained by dividing the absolute width of the bin by the sum of absolute widths over all bins. It is shown in equation 1,where relWidth(i) and absWidth(i) represent the relative and the absolute size of width at *i*th bin respectively.

$$relWidth(i)=absWidth(i)/\sum_{i}^{b_F} absWidth(i) \quad (1)$$

Similar to the process of hand extraction, the number of bins, b_F should be appropriately set. If the number of bins is too small, key features cannot be captured. An extreme case is using one bin. Although computationally less expensive, the pattern of width and the hand shape cannot be inferred from only the average hand width. The opposite extreme case is using the number of column pixels of the abstracted hand data as b_F. Although the shape of hand can be represented at the fine-grained level, it is computationally expensive. More importantly, by focusing on too specific and too local patterns of data, we may fail to extract more general data patterns. If there are more possible gestures such as stretching fingers. Each finger can have only one of two states as stretched or

folded, so that most of the commonly used hand gestures are combinations of the states of all five fingers. As a first step toward the recognition of the states of fingers, we attempt to recognize the number of fingers in each bin without distinguishing whether a finger is stretched or folded or by specifically identifying the fingers in each bin. To find the number of fingers within a bin, connected component analysis is used. Instead of filtering out the component with the smallest number of events as we did in the filtering process, the number of connected components for each bin is counted.

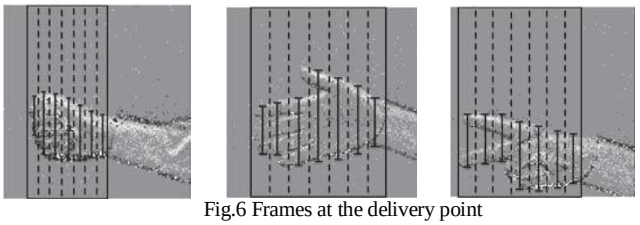


Fig.6 Frames at the delivery point

Next, as a step toward the recognition of the state of fingers, we attempt to recognize whether any of the fingers except the thumb is stretched or not. This could be easily recognized by comparing ratio of the length from the tip of fingers to the thumb to the length from the thumb to the wrist as shown in following fig. 7. This can roughly estimate the location of thumb by locating the column with the maximum width. This feature is called as Horizontal Ratio.

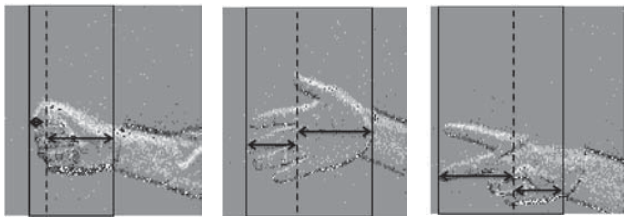


Fig.7 Horizontal Ratios [1]

E. Classification

Once the features are extracted, a machine learning algorithm is applied to build a model for prediction. In this paper, we use naive Bayes algorithm which is a simple version of Bayesian network with an assumption that all the features are independent given the class. This method is an effective and fast method for static hand gesture recognition. This method is based on classifying the different gestures according to geometric-based invariants which are obtained from image data after segmentation; thus, unlike many other recognition methods, this method is not dependent on skin color. Gestures are extracted from each frame of the video, with a static background. The Naïve Bayes classifier works on a simple, but comparatively intuitive concept. Once the features are extracted, a machine learning algorithm is applied to build a model for prediction. In this paper, we use naive Bayes algorithm which is a simple version of Bayesian network with an assumption that all the features are independent given the class. In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter. Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple. Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In spite of their naive design and apparently over-simplified assumptions, naive Bayes classifiers have worked quite well in many complex real-world situations. An advantage of the naive Bayes classifier is that it only requires a small amount of training data to estimate the parameters necessary for classification. In this system Width of palm and width of fingers as input to the Bayes technique and Output will be classifying Rock, Paper, and Scissors hand gestures. Requires small amount of training data for classification.Training data for hand gesture is class feature Rock, Paper and scissors. Training data for mouse free Interface are (class feature) is Right click ,Left click, mouse movement.(Class variable) Rock, Paper, Scissors

VII. CONCLUSION

This paper proposes a method to classify bare hand gesture using dynamic vision sensor (DVS) camera. Specifically, we focused on classifying three different throws delivered by a player playing rock-paper-scissors game. We first described the properties of DVS camera which only responds to the pixels with luminance changes. Then, we proposed a method to detect a delivery point where the final throw is made. Once the delivery point is detected, only the still image of hand at the delivery point is used for classification. The region of hand is first extracted by locating the wrist point where the changes of the width of hand is the biggest, and then we extract the distribution of width within the hand or the number of connected components for each segment as features. The experimental results were promising; using component-based method and ratio of the length of finger to the length of palm results in 89% of the accuracy. Component-based feature extraction method performed slightly better than width-based method, and the ratio of the length of finger to the length of palm could contribute to the enhancement of classification accuracy. However, there is a lot room for improvement. We discussed that the threshold for detecting delivery point should be adaptively changing depending on the size of actual object, and that the hand region should be more accurately detected in a situation where forearm is not at 90-degree angle at the upper body. It is expected that studies on detecting me finger state (i.e., open or folded) goes a step further. This work will be based on using the component-based method and it is expected to help make our system applicable in more general situations with numerous possible hind gestures. Furthermore, we plan to compare the classification accuracy and computational cost of using a DVS camera with those using conventional frame-based camera.

REFERENCES

[1] Eun Yeong Ann, Jun Haeng Lee, Tracy Mullen, John Yen,"Dynamic Vision Sensor Based Bare Hand Gesture Recognition",978-1-4244-9915-1/11/$26.00,2011 IEEE

[2] Pragati Garg, Naveen Aggarwal and Sanjeev Sofat," Vision Based Hand Gesture Recognition", World Academy of Science, Engineering and Technology 49, 2009

[3] C. A. Bouman," Connected Component Analysis", Digital Image Processing - January 9, 2012

[4] Preeta Rajamani," Best Practices in Gestural Design", Bentley University

[5] http://www.worldrps.com /game-basics

[6] Sanjay Meena,"A study on hand gesture recognition technique", Department of Electronics and Communication Engineering National Institute of Technology, 2011.

[7] Akhil Khare, "Hand gesture recognition based on collaborative augmentedreality environment for human –computer interaction",Journal of information ,knowledge and research in electronics and communication engineering,2011