# A Real Time Human Computer Interface Based on Gesture Recognition for Modification of Images

Chetan Aivalli, Prakash Biswagar, Vinaykumar S Lonimath

**Abstract--** This paper designs a simple, natural and real time system for gestural interaction between the user and computer for providing a dynamic user interface. This technique may also called as gesture recognition, Human computer interaction (HCI).The basis of our approach is a real-time tracking process to obtain the moving hand from the whole image, which is able to deal with a large number of hand gestures, and a recognition process that identifies the hand posture/gesture . The use of a visual memory (Stored database) allows the system to handle variations within a gesture and speed up the recognition process through the storage of different variables related to each gesture. The gesture recognition system uses image processing techniques for detection, tracking, and recognition on the basis of detection of some useful shape based features like orientation, area, centroid, extrema, location. In this paper we adopted Histogram based tracker for tracking which is based on CamShift algorithm and converting hand gestures in to a meaningful command. CamShift algorithm mainly used for tracking the objects but here we have used it for tracking and gesture recognition and we got high recognition rate compared to previous proposed methods in real time in a given environment condition. These image based features are extremely variant to different light conditions and other influences. So we have assumed that the background is static and has different color compared to hand for better results. To implement this approach we will use a simple web cam. The overall algorithm implemented in Matlab Tool. And results are observed in Picasa.

**Index Terms**--- Camshift, Gesture recognition, Histogram Based Tracker, Human Computer Interaction (HCI), Real Time Tracking.

## 1 INTRODUCTION

Human–computer Interaction (HCI) or sometimes called (MMI) Man Machine Inter action involves the study, planning, and design of the interaction between User and Computers. Many ways of communications are used between human and computer, while using gesture is considered to be one of the most natural ways in a virtual reality system. Hand gesture is one of the typical methods of non-verbal communication for human beings and we naturally use various gestures to express our own intentions in everyday life. Gesture recognition is a modern, new, innovative and current topic of engineering with the goal of interpreting human gestures via mathematical algorithms. Gestures can originate from any bodily motion or state but commonly originate from the face or hand. Gesture recognition enables humans to communicate with the machine (HMI) and interact naturally without any mechanical devices. Using the concept of gesture recognition, it is possible to point a finger at the computer screen so that the cursor will move accordingly. This could potentially make conventional input devices such as mouse, keyboards and even touch-screens redundant. It is bridge between machines And humans than primitive text user interfaces or even GUIs (Graphical user interfaces), which still limit the majority of input to keyboard and mouse.

Gesture recognition is useful for processing information from humans which is not conveyed through speech or type. As well, although the study of gesture is still in its infancy, some broad categories of gestures have been identified by researchers. A single emblematic gesture can have very different significance in different cultural contexts, ranging from complimentary to highly offensive The page List of gestures discusses emblematic gestures made with one hand, two hands, hand and other body parts, and body and facial gestures.

The video-based approaches claim to allow the signer to move freely without any instrumentation attached to the body. Trajectory, hand shape and hand locations are tracked and detected by a camera. By doing so, the signer is constrained to sign in a closed, some-how controlled environment. The amount of data that has to be processed to extract and track hands in the image also imposes a restriction on memory, speed and complexity on the computer equipment. Current focuses in the field include emotion recognition from the face and hand gesture recognition.

Here we proposed hand gesture recognition. Many approaches have been made using cameras and computer vision algorithms to interpret sign language. Flutter, a start-up in Palo Alto, California, is allowing anyone with a Mac or Windows computer and webcam to download an application that allows them to control music and video applications such as Spotify, iTunes, Windows Media Player, QuickTime, and


- *Chetan Aivalli is currently pursuing masters in communication systems in R.V. College of engineering Bangalore 560059, India, PH-7795215342. E-mail: chetanaivalli@gmail.com*
- *Prakash Biswagar is currently Associate Professor in Electronics and Communication department in R.V. College of engineering Bangalore 560059, India, PH-9740652188. E-mail: prakashbiswagar@rvce.edu.in*
- *Vinaykumar S Lonimath is currently pursuing masters in communication systems in R.V. College of engineering Bangalore 560059, India, PH-7411399124. E-mail:vinay3368@gmail.com*

VLC using gestures. Recently Samsung launched TV which can be controlled by gestures and even HTC using it in mobiles. A recognition system that we proposed comprises of a framework of two most important entity, which are preprocessing (image processing) and recognition classification it can be conducted with techniques from computer vision and image processing and developing some tracking algorithms and by using the basic concepts of image processing technology we can propose the gesture recognition system in real time. The literature includes ongoing work in the computer vision field on capturing gestures or more general human pose and movements by cameras connected to a computer.

The paper is further subdivided as follows: section 2 provides a gist of related work done for gesture recognition in the field of human computer interaction. Section 3 presents the system design for the proposed gesture recognition system. Section 4 details about the experimental results and analysis .The effort ends in Conclusion that is in the section 5 .The scope of present work and future scope of the present research effort has been stated in section 6.References used for designing the gesture recognition system are summarized in section 7.

## 2 Related Work

There were many gesture recognition techniques developed for tracking and recognizing various hand gestures. Each one of them has their own pros and cons. Initially the wired technology/electronic switches were used for interaction, but these require power to operate such as all wireless devices require power. The older one is wired technology, in which users need to tie up themselves with the help of wire in order to connect or interface with the computer system even user can not freely move in the room as they connected with the computer system via wire and limited with the length of wire. Instrumented gloves also called electronics gloves or data gloves is the example of wired technology. Even sound/speech can be given as input command but in real time there will be much noise that's why it not accepted widely and we need to de-noise it for it we need to implement many filters which will lead to high cost currently there are many techniques but they need to be operated online i.e. we should have the internet connection. We have pen computing technology which is (devices like stylus) used for the interface purpose in mobile communication. But we need to carry physical device always with us and it has lack of sensitivity.

Nowadays multi touch technology is running but it requires physical contact with the device. Here we implement such algorithm which can recognize the gesture and for particular gesture suitable action will take place.

A wired glove ("data glove" or "cyber glove") is an input device for human–computer interaction worn like a glove. These instrumented gloves made up of some sensors, provide the information related to hand location, finger position orientation etc through the use of sensors. These data gloves provide good results but they are extremely expensive to utilize in wide range of common application. Example "Project Digits" is a Microsoft Research Project. Project Digits is an input which can be mounted on the wrist of human hand and it captures and displays a complete 3D graphical representation of the user's hand on screen without using any external sensing device or hand covering material. But we always need to carry the IR equipment.

Various sensor technologies are used to capture physical data such as bending of fingers. Often a motion tracker, such as a magnetic tracking device or inertial tracking device, is attached to capture the global position/rotation data of the glove. These movements are then interpreted by the software that accompanies the glove, so any one movement can mean any number of things. Gestures can then be categorized into useful information, such as to recognize Sign Language or other symbolic functions. Expensive high-end wired gloves can also provide hepatic feedback, which is a simulation of the sense of touch. This allows a wired glove to also be used as an output device. Traditionally, wired gloves have only been available at a huge cost, with the finger bend sensors and the tracking device having to be bought separately

Data gloves are then replaced by optical markers. These optical markers project Infra-Red light and reflect this light on screen to provide the information about the location of hand or tips of fingers wherever the markers are wear on hand, the corresponding portion will display on the screen. These systems also provide the good result but require very complex configuration.

The Sixth Sense technology contains a pocket projector, a mirror and a camera contained in a pendant-like, wearable device. Both the projector the camera and sensors are connected to a mobile computing device in the user's pocket. The projector projects visual information enabling surfaces Sixth Sense prototypes cost approximately $350 million

The success of approach discussed in paper [1] for hand gesture recognition based on shape features is highly influenced by some constraints like hand should be straight for orientation detection in image, if it will not be followed then result could be unexpected or wrong and also we fix the new parameter to detect the presence of thumb. In paper [2], the approach is based on calculation of three combined features of hand shape which are compactness, area and radial distance. Compactness is the ratio of squared perimeter to area of the shape. If compactness of two hand shapes are equal then they would be classified as same, in this way this approach limits the number of gesture pattern that can be classified using these three shape based descriptors and only 10 different patterns have been recognized[1].

Accelerometers, data glove [3], sensors/actuators and other input devices used to capture the user movement and control the selection, manipulation and movements of objects in virtual scenes [4]. The use of these devices is hindered by numerous factors such as awkwardness, rigidity, and prone to distortion from the physical environment. These devices have cost which is prohibitive for their frequent use by the general user. Hence they are mostly used by highly trained professionals like the surgeons and ace pilots to train and accomplish their operations in the virtual environment.

Liu and Lovell [4], proposed an interesting technique for real time tracking of hand capturing gestures through a web camera and Intel Pentium based personal computer. The proposed technique is implemented without any use of sophisticated image processing algorithms and hardware. Atia et al. [6] designs a tilting interface for remote and quick interactions for controlling the directions . It uses coin sized 3D accelerometer sensor for manipulating the application.

A number of systems have been implemented that recognize various types of sign languages. For example, Sterner was able to recognize a distinct forty word lexicon consisting of pronouns, nouns, verbs, and adjectives taken from American Sign Language with accuracies of over 90%. A system developed by Kadous recognized a 95 word lexicon From Australian Sign Language and achieved accuracies of Approximately 80%.Murakami and Taguchi were able to adequately recognize 42 finger-alphabet symbols and 10 sign-language words taken from Japanese Sign Language [8]. Lu [10] and Matuso [7] have also developed systems for recognizing a subset of Japanese Sign Language. Finally, Takahashi and Kishino were able to accurately recognize 34 hand gestures used in the Japanese Kana Manual Alphabet [9].

The major drawbacks of such techniques are they are very complex and highly sophisticated for developing an actionable procedure to make the necessary jigs and tools for any typical application scenarios. This problem can be overcome by pattern recognition methods having lower hardware and computational overhead.

In [12] Zhou (2005) he did articulated object ( body/hand posture) recognition. He approached this by inberted indexing an image using local image features the speed was about 3s/query, the main drawback its Non-real time. In [13] Derpnis proposed technique for hand gesture recognition, he implemented it by using the molecular temporal sequence of images technique. Its comes in a category appearance based syntactic, the speed was about 8s/frame the main drawback its Non-real time. In [14] Lin 2004 tracked the articulated hand motion in a video sequence in this he searched for optimal motion estimate in a high dimentional configuration space, which comes under the category hand model based statistical the speed was about 2s/frames the main drawback its Non-real time.

In [15] Ye 2004 comes with the new technique classify manipulative and controlling gestures in which they used 3D hand appearance using a region based coarse stereo matching algorith ie its depends on the Stereo vision statistiical the advantage is that its real time but two camaras calibration required which is very complex.

In [16] Kolsch 2004 did fast tracking for non-rigid and highly articulated object such as hands he used flock of KLT features/colour to faciliate 2D hand tracking and posture recognition from a monocular view,its appearance based statistical and real time but strict requirement on hand pose configuration for recognition required.

In [11] Barczak 2005 proposed real time hand tracking he approached this technique by the Viola-Jones method which comes under the category appearance based statistics. The advantage is that real-time but limited to single posture. 3D hand model- rich description and wide class of hand gestures. However, as the 3D hand model is a complex articulated deformable object with many degrees of freedom, a very large image database is required to cover all the characteristic hand images under different views. Matching the query images from the video input with all hand images in the database is time-consuming and computationally expensive. Another limitation of 3D hand model-based approaches is the lack of the capability to deal with singularities that arise from ambiguous views. Big issue is the scalability problem, where a 3D hand model with specific kinematic parameters cannot deal with a wide variety of hand sizes from different people

## 3. System Design

The real time tracking here we have used the histogram based back projection. It tracks an object by using the Continuously Adaptive Mean Shift (CamShift) algorithm. The principle of the CamShift algorithm is based on the principles of the algorithm Meanshift which provides accurate localization and efficient matching without expensive exhaustive search. It provides reliable way of tracking objects whose appearance is defined by color and doesn't change color over time. It's not only limited to color, it could be use edge, texture etc. Camshift is able to handle the dynamic distribution by adjusting the size of the search window for the next frame based on the moment zero of the current distribution of images. In contrast to the algorithm Meanshift who is conceived for the static distributions, Camshift is conceived for a dynamic evolution of the distributions so we adopted CamShift in our system. The algorithm that tracks hand movement using a one-dimensional histogram consisting of quantized channels from the HSV color space. Since the algorithm is designed to consume the lowest number of CPU cycles possible, a single channel (hue) is considered in the color model. This heuristic is based on the assumption that flesh color has the same value of hue. Furthermore, a bandwidth of acceptable color values is defined to allow the tracker to compute the probability that any given pixel value corresponds to flesh color.

### 3.1 Histogram back projection and Color probability distribution.

In order CAMSHIFT can track colored object, it needs a probability distribution image. They use the HSV color system and using only hue component to make the object's color 1D histogram. This histogram is stored to convert next frames into corresponding probability of the object. The probability distribution image itself is made by back projecting the 1D hue histogram to the hue image of the frame. The result called back project image. CAMSHIFT is then used to track the object based on this back project image. Regarding histogram back projection, it is a technique to find the probability of a histogram in an image. It means each pixel of the image is evaluated on how much probability it has to the histogram. The probability distribution image (PDF) may be determined using any method that associates a pixel value with a probability that the given pixel belongs to the target. A common method is known as Histogram Back-Projection. In order to generate the PDF, an initial histogram is computed at Step 1 of the CamShift algorithm from the initial ROI of the filtered image. The back-projection of the target histogram with any consecutive frame generates a probability image where the value of each pixel characterizes probability that the input pixel belongs to the histogram that was used. The histogram bin values are rescaled to the new range [0, 255], where pixels with the highest probability of being in the sample histogram will map as visible intensities in the 2D histogram back-projection image. [17]

### 3.2 Mass center, orientation and scaling calculation

The mean location of a target object is found by computing zero[th], first and second order image moments and first image moments. The orientation and scale are calculated using second image moments.

$$M_{00} = \sum_x \sum_y P(x,y) \quad (1)$$

$$M_{10} = \sum_x \sum_y xP(x,y) \quad (2)$$

$$M_{01} = \sum_x \sum_y yP(x,y) \quad (3)$$

$$M_{20} = \sum_x \sum_y x^2P(x,y) \quad (4)$$

$$M_{02} = \sum_x \sum_y y^2P(x,y) \quad (5)$$

Where $P(x, y) = h(I(x, y))$ is the back projected probability distribution at position x, y within the search window I(x, y) that is computed from the histogram *h* of *I*. The target object's mean position can then be computed with

$$y_c = M_{01} / M_{00} \quad (6)$$

$$x_c = M_{10} / M_{00} \quad (7)$$

While its aspect ratio

$$Ratio = \frac{M_{20}}{x_c^2} \Big/ \frac{M_{02}}{y_c^2} \quad (8)$$

Is used for updating the search window with

$$Width = (\sqrt{2M_{00}})\ ratio \quad (9)$$

$$Height = \sqrt{2M_{00}} \Big/ ratio \quad (10)$$

The position and dimensions of the search window are updated iteratively until convergence. The whole tracking system and recognition is explained in fallowing steps.

A] Initially we trigger the webcam manually and make it to take the video
B] After initial adjustments we apply CamShift algorithms on it.
C] The tracking algorithm can be summarizing with these steps and shown in fig 1.

1) Initially we choose the region of interest, which contains the object to be tracked.
2) Make a color histogram of that region as the object model.

3) We make a probability distribution of the frame using the color histogram. We fallowed the histogram back projection method.
4) Based on the probability distribution image, find the center mass of the search window using mean-shift method.
5) Center the search window to the point taken from step 4 and iterate step 4 until convergence.
6) Process the next frame with the search window position from the step 5.

D] Once tracking is done efficiently then store gestures in database and by giving some threshold value or logical value. We can define no of gestures and passing those gestures into suitable commands is done by giving some threshold values. In Image browser Picasa we can perform the desired action, by using conditional statements in code.

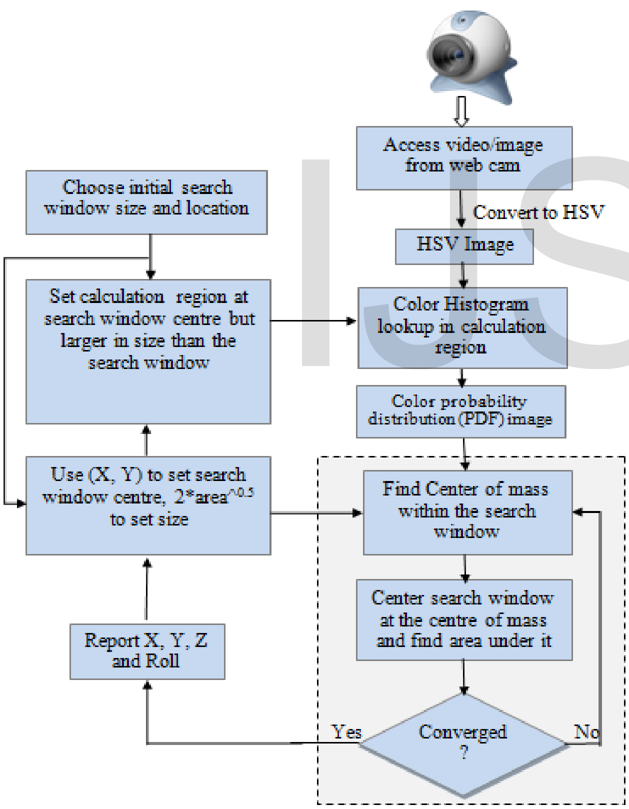


Figure 1: Block diagram of the tracking system

The detailed block and functional diagram is as shown in the fig 1.For CamShift tracking we need to convert RGB image to HSV these are done in initial adjustments.

## 4 Results and Simulation

The whole proposed method is developed in Matlab code and modification of images is done in Picasa image browser. The hardware requirements are Minimum Intel Pentium IV Processor, 1 GB RAM, and 20 GB HDD, CD-ROM, Web cam 720 pixels, Picasa or any one image browser.

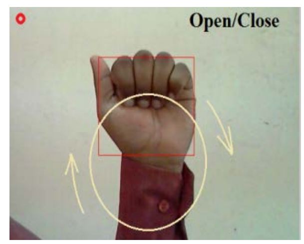


Fig: 2.1 Open/Close

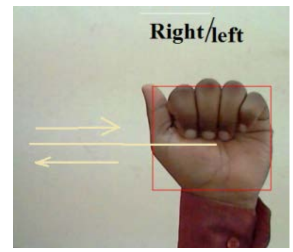


Fig: 2.2 Right/Left

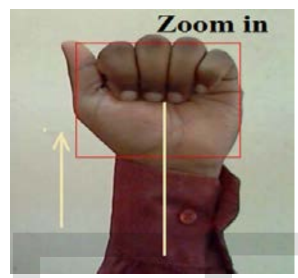


Fig: 2.3 Zoom in

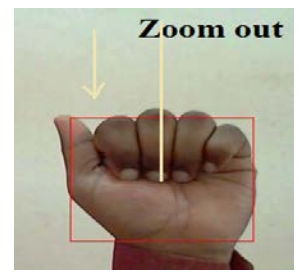


Fig: 2.4 Zoom Out

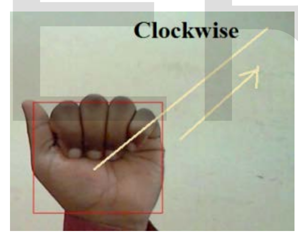


Fig: 2.5 Clockwise

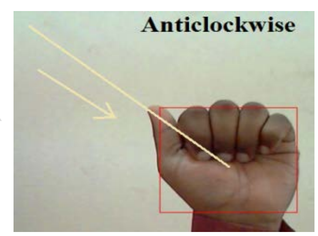


Fig: 2.6 Anticlockwise

Figure 2: Various types of gestures defined for the proposed system

In Figure 2 we have showed the various types of gestures defined for our system .these gestures are recognized by the algorithms which we have proposed and corresponding action will be taken place these gestures are taken as command and these commands are passed to the Picasa Image Browser for particular valid gesture the particular action will takes place. In fig 3 we have showed the real time results in Picasa. After passing the commands to Picasa fallowing actions will be taken place without much delay.

Fig 3.1 shows array of images. Out of these we select any one image by gesture shown in fig 2.1 and the selected image shown in fig 3.2 the same gesture will be used for closing purpose also by defining flag. Similarly for gesture (fig 2.2)left/right the results will be next/previous image as shown in fig 3.3 and 3.4.For zoom in and Zoom out gestures (fig2.3,2.4) the results shown in fig 3.5 and 3.6. In the same way for Clockwise and anticlockwise gestures (fig 2.5 and 2.6) the results shown in fig 3.7 and 3.8

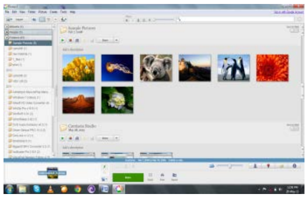
Fig:3.1 Picasa image browser


Fig:3.2 Open/Close

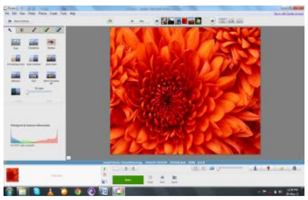
Fig:3.3:Next Image

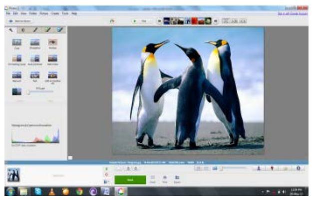
Fig:3.4 Previous Image


Fig:3.5 Zoom in

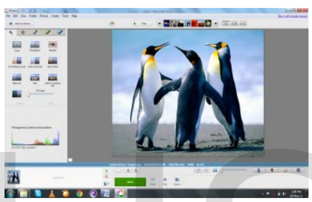
Fig:3.6 Zoom Out

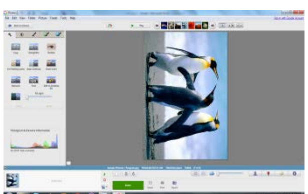
Fig:3.7 Clockwise Rotate

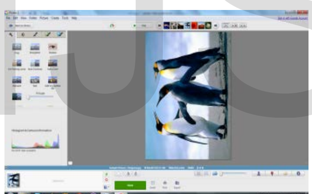
Fig:3.8 Anticlockwise Rotate

Fig:3 Modification on the images based on the predefined gestures

After many times executing this code we have plotted the recognition rate which is shown in fig 4. As we can see in the fig 4 we have recognition efficiency 78%, 86%, 87%, 79%, 76%, 88%,86%,79%,for open, Next image, previous image, Rotate clockwise, Rotate anticlockwise, zoom in, zoom out and close respectively. Recognition rate bit low for clockwise and anticlockwise, open and close it's due to gesture complexity by adopting easy gestures we can reach much more good value as we can see for next and previous image. And even we have considerable best values for zoom in and zoom out.

Thus by defining various different gestures we can track them using Camshift and recognise them.Taking those as commands and passing them to image browser we can do modification on images in real time by this we can conclude that by gesture recognition we can control not only Picasa but any system .

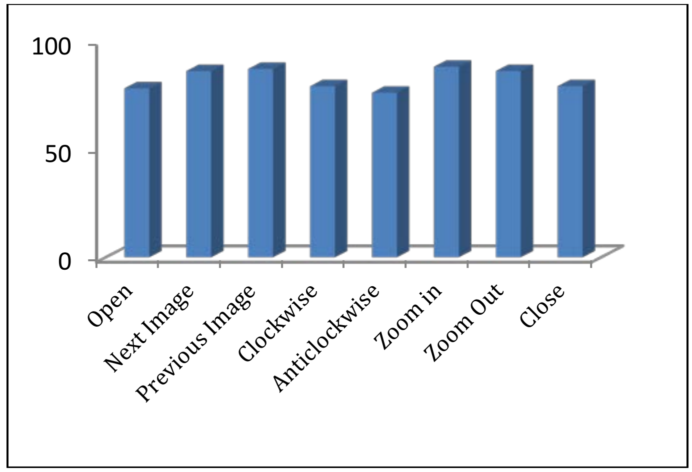


Fig: 4 Graphical representation of Gesture Recognition rate

## 5. Conclusion

In this paper we have proposed a real time gesture recognition technique or HCI based on histogram based tracker with less delay and high recognition rate in a given environment and the proposed algorithm is best suited for real time application. We implemented this technique which is used for modification of images in the image browser Picasa and gives faithful and real time results with high efficiency. By this system we can control not only Picasa but any software. The proposed research work can be implemented in various domains like education field which has wide area and even in playing computer games.

## 6. Future Work

The proposed system is purely real time and has very good recognition rate in static background and for a given environment condition. As in dynamic background and change in light conditions we get low recognition rate, we can overcome from these challenges by developing new algorithms or by improving existing algorithms. Our future research will address more complex gestures like by using two hands with much higher recognition rate and dynamic background with different light conditions. And even we can develop systems like kinect.

## Acknowledgment

The authors wish to thank Dr. S Ravi Shankar and Dr. Hariprasad S.A Professors of Department of Electronics and Communication engineering R.V. College of engineering Bangalore for their moral support.

## REFERENCES

[1]Meenakshi panwar and Singh Mehra "Hand Gesture Recognition for Human Computer Interaction ",in proceeding of IEEE International Conference On Image Information Processing (ICIIp 2011), Waknaghat ,India, November 2011

[2] Amornched Jinda-apiraksa, Warong Pongstiensak, and Toshiaki Kondo, "A Simple Shape-Based Approach to Hand Gesture Recognition", in Proceedings of IEEE International Conference on Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), Pathumthani, Thailand , pages 851-855, May2011""

[3] N. Karlsson, B. Karlsson and P. Wide, "A glove equipped with finger flexion sensors as a command generator used in a fuzzy control system," *IEEE Trans. On Instrumentation and measurement*, pp. 1330-1334, 1998. [5] D. W. F. V. Krevelen and R. Poelman, "A Survey of Augmented Reality Technologies, applications and limitation," The International Journal of Virtual Reality, vol. 9 9(2), pp. 1-20, 2010.

[4] N. Karlsson, B. Karlsson and P. Wide, "A glove equipped with finger flexion sensors as a command generator used in a fuzzy control system," *IEEE Trans. On Instrumentation and measurement, pp. 1330-1334, 1998.*

[5] N. Liu and B. Lovell, "Mmx-accelerated Realtime Hand Tracking System," *In Proceedings of IVCNZ*, 2001.

[6] Ayman Atia and Jiro Tanaka, "Interaction with Tilting Gestures in Ubiquitous Environments," In International Journal of UbiComp (IJU), Vol.1, No.3, 2010.

[7]R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," *In Proceedings of ICIP02*, pp. 900-903, 2002.

[8] C.H, Messom and A.L.C. Barczak, Fast and Efficient Rotated Haar-like Features Using Rotated Integral Images", *In Proceedings of Australian Conference on Robotics and Automation (ACRA2006)*, pp. 1-6, 2006.

[9] G. R. Bradski. Computer video faces tracking for use in a perceptual user interface. Intel Technology Journal, Q2, pp. 1-15,1998.

[10] N. Conci, P. Cerseato and F. G. B. De Natale, "Natural Human-Machine Interface using an Interactive Virtual Blackboard," *In Proceeding of ICIP 2007, pp. 181-184, 2007.*

[11] A. L. C. Barczak, F. Dadgostar, and M. Johnson, "Real-time hand tracking using the Viola and Jones method," in Proc. International Conference on Image and Signal Processing, 2005, pp. 336–441.

[12] H. Zhou and T. Huang, "Okapi-Chamfer matching for articulate object recognize- tin," in Proc. International Conference on Computer Vision, 2005, pp. 1026–1033.

[13] K. G. Derpanis, R. P. Wildes, and J. K. Tsotsos, "Hand gesture recognition within a linguistics-based framework," in Proc. European Conference on Computer Vision 2004, pp. 282–296.

[14] J. Y. Lin, Y. Wu, and T. S. Huang, "3D model-based hand tracking using stochastic direct search method," in Proc. IEEE International Conference on Automatic Face and Gesture Recognition, 2004, pp. 693–698.

[15] G. Yes, J. J. Corso, and G. D. Hager, "Gesture recognition using 3d appearance and motion features," in Proc. CVPR Workshop on Real-Time Vision for Human- Computer Interaction, 2004, pp. 160–166.

[16] M. Kolsch and M. Turk, "Fast 2D hand tracking with flocks of features and multi- cue integration," in Proc. Conference on Computer Vision and Pattern Recognition Workshop, vol. 10, 2004, pp. 158–165.

[17] Gary R. Bradski, Microcomputer Research Lab, Santa Clara, CA, Intel Corporation "Computer Vision Face Tracking for Use in a Perceptual User Interface"