Real Time Hand Gesture Recognition for Smart Classroom Environment

Swapnil S. Potnis[#1], Asst. Prof. Aditi S. Jahagirdar[#2]

[#1] *Student, Dept. of Information Technology, MIT College of Engg., India*
[#2] *Asst. Professor, Dept. of Information Technology, MIT College of Engg., India*

***Abstract*— In today's expeditiously changing world, the physical classrooms are soon getting replaced by virtual classrooms. In such an environment, it is important to capture the raising-of-hand gesture of the participants and convey it to the lecturer. Gesture Recognition is one of the optimal techniques which help us capture such an event. The existing systems make's user wear colored gloves which is not a user-friendly approach. Some require costly in-depth sensor camera's for using them. We propose a systematic approach to detect face using cascading classifiers, recognize the hand-gesture and annotate the video. This is made possible using webcam available at affordable prices. The results obtained show that the system can work under a varying degree of background lightning condition as well as illumination. The location and the hand orientation in accordance with the face detected help us in achieving greater efficiency..**

***Keywords*— Raising-Hand, Colored Gloves, User-Friendly, Cascade Classifiers, Hand-Gesture, Background Lighting, Illumination.**

I. INTRODUCTION

Computers are used by many people either at their work or in their spare-time. Special input and output devices have been designed over the years with the purpose of easing the communication between computers and humans, the two most known are the keyboard and mouse [15]. As the complexities of human needs have turned into many folds and continues to grow so, the need for Complex programming ability and intuitiveness are critical attributes of computer programmers to survive in a competitive environment. However, a shift towards a user friendly environment has driven them to revisit the focus area.

One of the attractive methods for providing natural human-computer interaction is the use of the hand as an input device rather than the cumbersome devices such as keyboards and mouse, which requires the user to be located in a specific location to use these devices. The most important thing in hand gesture recognition system is the input features, and the selection of good features representation. Keeping track of the hand from a sequence of images helps achieve better results during the detection and recognition of the gesture. This helps us in achieving accurate results in less time.

Hand Gesture Recognition is the solution for a natural human – computer interface. These hand gestures could be of two type's viz., static and dynamic [5]. The static hand gesture is a particular hand configuration and position while the dynamic hand gesture is that of a moving gesture interpreted by a sequence of images. The main focus of this paper is dynamic hand gesture-recognition

II. RELATED WORK

Gesture recognition is a challenging task that is often addressed with complex sensors and methods, such as the use of depth sensors and multiple classifier systems, under controlled acquisition conditions.

A. Real-time hand gesture recognition using a Colored Glove

The recognizer is formed by three modules: The first module, fed by the frame acquired by a webcam, identifies the hand image in the scene. The second module, a feature extractor, represents the image by a nine dimensional feature vector. The third module, the classifier, is performed by means of Learning Vector Quantization. The recognizer, tested on a dataset of 907 hand gestures, has shown very high recognition rate.

Fig. 1 Hand Gesture using Colored Gloves

The main disadvantage of this technique is that the user-friendliness of the whole process of hand detection is compromised as the user has to wear the gloves every time that he wishes to use the system.

B. Hand gesture interaction using Color-Based Method for Table top

A hand gesture recognition method based on color marker detection is presented where four types of colored markers (red, blue, yellow and green) are mounted on the two hands. SixthSense is a wearable gestural interface that augments the physical world around us with digital information and lets us use natural hand gestures to interact with that information. But using of colored markers lead to reduction in usability of the system as the user has to wear the colored markers every time and for every motion.

C. Sensors for Gesture Recognition Systems

The gesture recognition system (GRS) consisting of multiple cameras along sensors in-built in them. In a GRS, the gesturer is embedded within the context of use (the environment and gestures). The input to the capture sensor is a physical stimulus (e.g., light waves, soundwaves, etc.) which is modulated by the gesture. The physical realization of the sensor is the sensor platform. Determination of the sensor stimulus, context of use, and sensor platform are major preliminary design issues in GRSs. Thus, these three components of the GRS form the basic structure of our taxonomy. The range of the sensor unit may vary with respect to the human being and other moving objects with one sensor giving a varying output as compared to the other.

D. Skin Detection in Luminance Images using Threshold Technique

An experimental investigation, which aims at using different threshold values to classify the human skin. The chromatic skin colour model was used to model skin colour in the r-g chromaticity space. This model based on testing database containing many of human images. It has been found that the threshold value is an important tool for increasing the human skin regions detected in colour images containing luminance. The skin colour pixels are white and the other ones are black in skin-segmented image, the threshold value used in skin segmented image is called an optimal threshold value, where the minimum increase in region size is observed while decreasing the threshold value.

E. Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques

The work consisted of a novel and real-time system for interaction with an application or videogame via hand gestures. The system included detecting and tracking bare hand in cluttered background using skin detection and hand posture contour comparison algorithm after face subtraction, recognizing hand gestures.

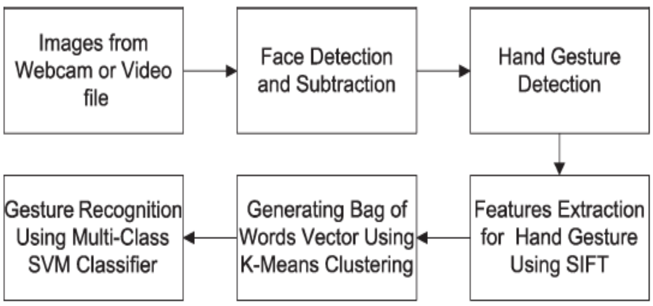

Fig. 2 Existing Work Architecture

A vector quantization technique helps in mapping keypoints from the entire available training image into a pack of unified dimensional histogram vector (bag-of-words) after K-means clustering. This histogram is then further taken as an input vector for a multiclass SVM in order to build the training classifier. In the testing stage, for every frame captured from a webcam, the hand is detected, the key-points are extracted for every small image that contains the detected hand gesture only and fed into the cluster model to map them into a bag-of-words vector, which finally recognizes the hand gesture.

III. PROPOSED WORK

In order the make the hand detection and recognition faster for a virtual classroom specific environment, the proposed work consists of using HAAR Cascade Classifier for accuracy along with Blob Detection for improved performance.

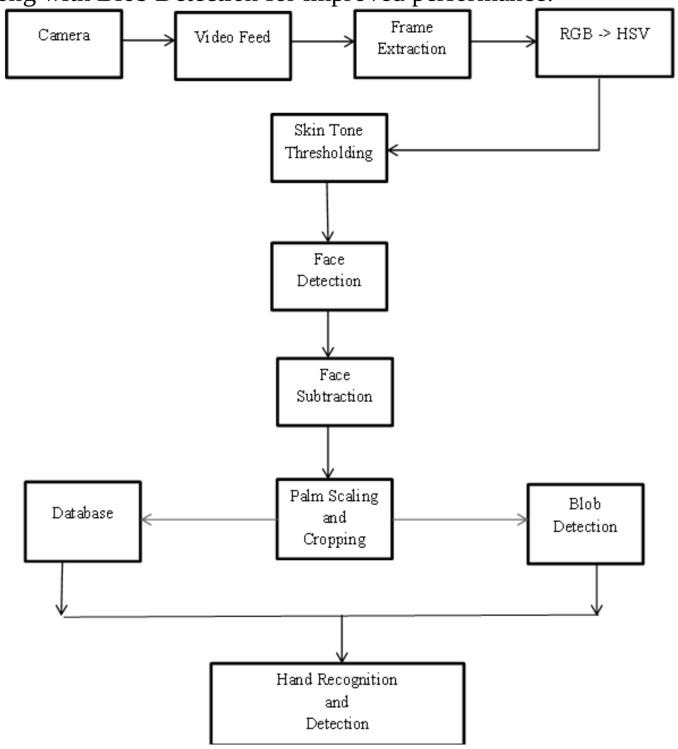

Fig. 3 Hand Recognition Workflow

A. Camera Feed

For capturing the video feed from the camera, Java library namely JMyron library is used. JMyron (aka WebcamXtra) is an external library for Processing that allows image manipulation without having to hard code everything.

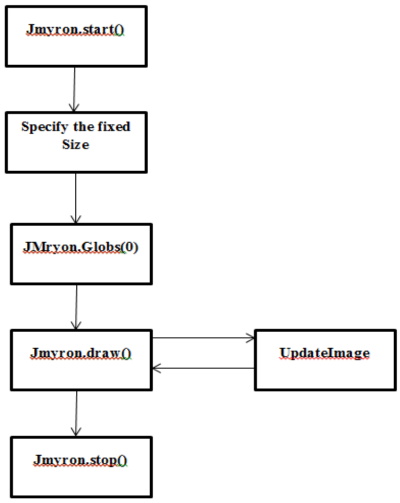

Fig.4 JMyron Architecture

Firstly we call the inbuilt function of JMyron named start() function to start image capture and its stop() function to stop. Always keep in mind to call JMyron. stop() when done in order to be call start() method the next time we need to. It's a good programming to call JMyron. stop() at the end of applet's stop() function, so that we'll stop capturing after we quit.

B. RGB -> HSV Colour Model

By default, an image is represented in an RGB format, meaning that for each pixel in the image each of the colours Red, Green and Blue has a value stating the intensity of that colour in the pixel. Since different people have different hand colours (e.g. Black hands vs. white hands) we cannot rely on the intensity of any one of the colours in order to extract the hand and its features. Therefore the RGB representation is converted to a Hue-Saturation-Value (HSV) representation, meaning that now for each pixel in the image there are three values – Hue, Saturation and Value.

- **Hue Calculation**

$$H=\begin{cases} 60^\circ * \left(\frac{G'-B'}{\Delta} mod6\right), Cmax = R' \\ 60^\circ * \left(\frac{B'-R'}{\Delta} + 2\right), Cmax = G' \\ 60^\circ * \left(\frac{R'-G'}{\Delta} + 4\right), Cmax = B' \end{cases} \quad ....(1)$$

- **Saturation Calculation**

$$S = \begin{cases} 0, \Delta = 0 \\ \frac{\Delta}{Cmax}, \Delta <> 0 \end{cases} \quad .... (2)$$

- **Value Calculation**

$$V = Cmax \quad ....(3)$$

By looking on different hand pictures and measuring their HSV values we concluded that we can find hue and saturation thresholds that have high correlation with typical hand colours. It has been found that typical skin HSV values are-($0.12 \leqslant S \leqslant 0.4$), ($0.94 \leqslant H \leqslant 1$) or ($0 \leqslant H \leqslant 0.137$).

C. Skin Tone Thresholding

Thresholding enables you to select ranges of pixel values in grey scale and colour images that separate the objects under consideration from the background. Thresholding converts an image into a binary image, with pixel values of 0 or 1. This process works by setting to 1 all pixels whose value falls within a certain range, called the threshold interval, and setting all other pixel values in the image to 0.

Skin Tone Thresholding can be brought about in the following two ways:-

a. Range:-

In this technique, the specific range of the user skin-tone range is pre-specified and system is made to detect all pixelates ranging in that range during the time of glob detection (to be discussed later).

b. Pixel Selection

In this type of Thresholding, the pixel range is specified by providing the maximum exposed skin-tone colour and the least specified skin-tone colour.

D. Face Detection

After the captured image is converted from RGB -> HSV and then applying the skin tone Thresholding to the converted image, it is then further sent to face detection module for face-detection and subtraction purpose. The purpose of face detection and subtraction is mainly because the skin tones of the face as well as the hand lie in the same threshold region. Hence in order to differentiate the hand from the face during the time of hand recognition, the face detection and subtraction is done first.

In order to make the face detection possible, HAAR Cascade Classifiers using OpenCV are used. The HAAR Cascade Classifier of the OpenCV makes use of the training set of images already feed to it from the very beginning and hence brings about the face detection. The co-ordinates of the detected face are then sent back to the system by the classifier for any further processing. Hence after the subtraction of the face from the image, now the hand detection takes place.

E. Blob Detection

In the area of image processing, blob detection refers to application of mathematical equations that are focused at detecting those regions of a digital image that differ in properties, such as brightness or colour of the region as compared to areas/portions surrounding those regions. In other words, a blob is that region of a digital image in which some attributes are constant while some may vary within a prescribed range of values; concluding in a way that all the points in a blob can be said to be similar to one another.

A blob as in its definition is an area which touches the pixels coming under the same logical state. The group of pixels in an image which belong to a particular blob are in a foreground state while the other pixels belong to the background state. Thinking of an image in binary terms, pixels belonging to the background are labelled as zero while every nonzero pixel is tagged along as a binary object.

On further proceeding, that which is easily identifiable by the any normal persons eye as several distinct but touching blobs may be calculated by the software as a single blob. Furthermore, any portion of a blob is considered to be a background which is in the background pixel state due to the lighting or reflection at the time of analysis. Blob detection is an algorithm used to determine if a group of connecting pixels are related to each other. This is useful for identifying separate objects in a scene, or counting the number of objects in a scene.

Algorithm for Blob Detection:

Step 1: Go through each pixel in the array

Step 2: If the pixel is a blob colour, label ='1'
Otherwise label=0

Step 3: Go to the next pixel - If it is also a blob colour and if it is adjacent to blob 1 then label = '1'
Else label = '2' (or more)

Step 4: Repeat Step 1 until all pixels are done

The histograms we use are in the HSV colour space since lighting and camera white balance changes cause variation largely in the saturation and value components with less effect on hue. For this reason, we only use the hue histogram in tracking and we threshold the saturation and value channels to try to maximize the presence of the tracked hue in the back projection. When the histogram is updated, the saturation and value threshold values are updated automatically using a binary search of the respective spaces to find the best back projection, the one containing the most positive pixels in the tracked area. This update assists with not only with lighting changes but also with quick movement by ensuring the track is not lost if the hand blurs with fast motion causing a slightly different histogram.

F. *Raised Hand Detection*

Raising hand is one of the most important types of interaction between students and lecturers in classroom. When an automatic system can be installed in virtual classroom to figure out which students raise their hands, it is possible to design more advanced applications for other education goals. The most important issue to improve this work is to design a good method to well segment the foreground subjects in the cluttered background. The camera feed is captured using JMyron library and per image captured is then processed for hand detection.

The conversion of RGB -> HSV onto the image then takes place and is further sent to HAAR Cascade Classifier using OpenCV for detecting the face and obtaining the co-ordinates of the detected face. After obtaining the co-ordinates of the face from the image is obtained, the image is then sent for blob detection specifying the skin-tone Thresholding value wherever the head is subtracted during the time of hand detection. The detected hand is indicated using a pointer and hence traces the movement of the hand which is made possible making use of blob detection.

IV. PERFORMANCE ANALYSIS

A. *Comparison Analysis*

For performance and analysis of the proposed work, papers [10], [58]–[63] that had a real time performance and discussed their recognition time and accuracy to compare their performance with our approach.

TABLE I
COMPARISON WITH EXISTING SYSTEMS

Paper	R.T	F.R	Background	M.U (MB)	Illu.lux	No. of Tests
[30]	0.030	320x240	White Wall	NA	NA	100
[31]	NA	100x100	Wall	NA	NA	57-76
[34]	0.13	640x480	Different	NA	NA	130
[33]	0.09 – 0.011	320x240	Cluttered	NA	NA	195
[1]	0.017	640x480	Cluttered	NA	NA	1000
[35]	NA	NA	NA	16	NA	NA
Our Method	0.013 -0.016	640x480	Cluttered	20.156	29 and above	50-60

- In [32], the hand gesture was detected using skin colour approach. Features for all the detected hand gestures used were extracted based on HaarWavelet Representation and stored in a database. During the recognition process, a measurement metric was utilized to measure the similarity between the features of a test image and those in the database.
- In [33], hand detection with Adaboost was used to trigger tracking and recognition. Then, adaptive hand segmentation was executed during detection and tracking with motion and colour cues. Finally, scale-space feature detection was applied to find palm-like and finger-like structures. Hand gesture type was determined by palm–finger configuration.
- In [34], first hand gesture was detected based on Viola–Jones method. Then, the Hu invariant vectors of the detected hand gesture were extracted, and a SVM classifier was trained for final recognition.
- In [30], the posture was detected using Haar-like features and the AdaBoost classifier. Based on the cascade classifiers, a parallel cascade structure was implemented to classify different hand postures.
- In [35], it acquired and pre-processed video image frame from camera, then extracted the normalized moment of inertia features and Hu invariant moments of gestures to constitute feature vector, which is inputted into SVM to achieved classification results.
- In [31], a neural network model was used to recognize a hand posture in an image. A space discretization based on face location and body anthropometry was used to segment hand postures.
- In [35], the heap memory consumption by the system is given which is compared with the memory consumption by the propose system. Even though memory utilization is greater as to the paper but performance is at par when compared to the performance results of the system stated in the paper.

V. RESULTS

A. *Time Comparison*

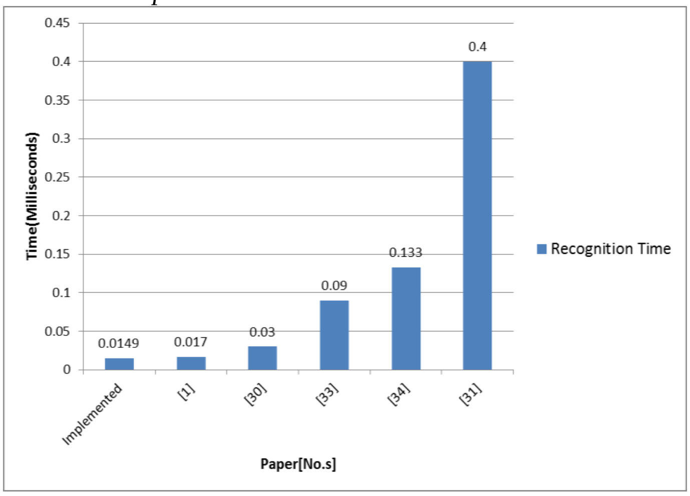

Fig.5 Time Comparison Graph

The graph shows the time taken by each system to recognize the hand gesture. The graph is plotted with the time taken in milliseconds and the system proposed in the respective papers. As seen from the graph, the proposed work takes little less time as compared to its existing predecessor by a few milliseconds

B. Performance With/Without Blur

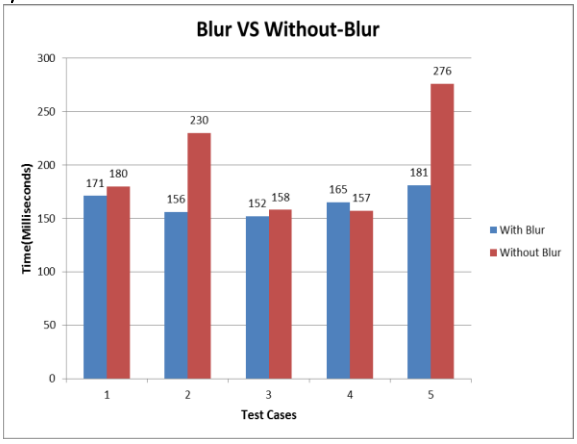

Fig. 6 Performance vs Time

The graph shows the performance of the system for different users where the Y axis is marked with the time taken for recognition of the hand and the X axis marks the different users that were involved for the analysis.

Also the two bars along the X axis stand for the performance of the individual user for Blur – On / Off. Blur helps in the reduction of the noise from the image hence enhancing the image for processing. More clear the image, better the performance of the system.

C. Background Variation

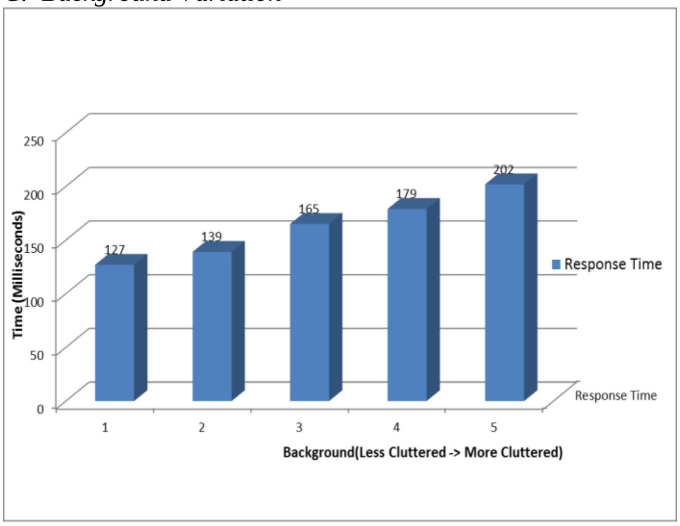

Fig. 7 Background Variation vs Time Graph

The above graph marks the response time of the system for various different backgrounds. The X axis consists of different backgrounds with their cluttered ness increasing in the ascending order of their number. The Y axis marks the respective time response for the corresponding background.

As is clear from the graph, the more the cluttered ness of the system the greater is the time taken by the system for the hand recognition. All title and author details must be in single-column format and must be centred.

D. Lightning Conditions

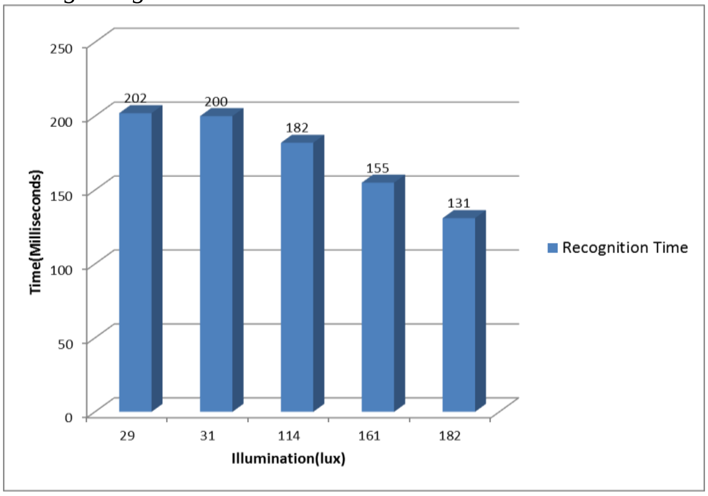

Fig. 8 Lightning Conditions vs Time Graph

The above graph stands for the response time of the system for the various Lightning Variations. The X axis denotes the illuminations in mux units while the Y axis denotes the response time for the same in milliseconds. The graph has been drawn considering a single user using the system for all the various lightning conditions.

E. Distance from Camera

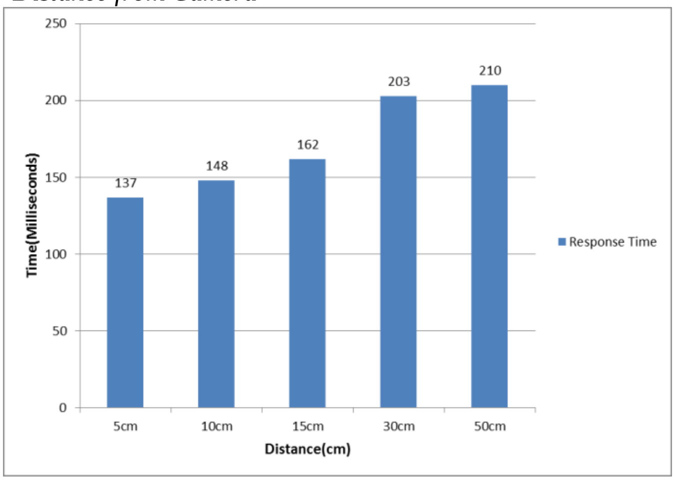

Fig. 9 Distance vs Time Graph

The above graph stands for the response time of the system for the distance of the hand from the camera. The X axis denotes the distance in centimetres while the Y axis denotes the response time for the same in milliseconds. The graph has been drawn considering a single user using the system for all the various distances.

VI. Conclusion

Hence a real-time hand gesture recognition system has been developed for a virtual classroom environment. The working functionality of the system is to detect the raise of hand gesture in the virtual classroom by the attendees. This is made possible by making use of HAAR Cascade Classifier for Face Detection and then applies Blob Detection for the filtered image. The detected hand is denoted by the pointer which follows the motion of the hand.

The developed system is at par as compared to the similarly existing systems. Also the system is capable of detecting the hand gesture in a cluttered background. The system also works fine in an environment where the lightning condition is as low as 29mux. The system has shown excellent results for a wide range of users varying in skin-tone values.

REFERENCES

[1] S Nasser H. Dardas and Nicolas D. Georganas, Fellow, IEEE," Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques", IEEE Transactions on Instrumentations and Measurements, Vol. 60, No.11, Nov. 2011

[2] Zhou Ren Junsong Yuan Nanyang, Singapore 639798 {renzhou, jsyuan}@ntu.edu.sg,"Robust Hand Gesture Recognition Based on Finger-Earth Movers Distance with Commodity Depth Camera ",MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA,Copyright 2011 ACM 978-1-4503-0616-4/11/11

[3] Qin, S., Zhu,, "Real-time hand gesture recognition from depth images using convex shape decomposition method", Journal of Signal Processing Systems pp. 1{12 (2013)

[4] B. Burger •I. Ferrané •F. Lerasle •G. Infantes," Two-handed gesture recognition and fusion with speech to command a robot, Auton Robot DOI 10.1007/s10514-011-9263-y.

[5] Siddharth S. Rautaray • Anupam Agrawal," Vision based hand gesture recognition for human computer interaction: a survey", Artif Intell Rev DOI 10.1007/s10462-012-9356.

[6] Sushmita Mitra, Senior Member, IEEE, and Tinku Acharya, Senior Member, IEEE," Gesture Recognition: A Survey", IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews, Vol. 37, No. 3, May 2007

[7] S.M. Shitole, S.B. Patil, "Dynamic Hand Gesture Recognition using PCA, Pruning and ANN", International Journal of Computer Applications , Volume 74– No.2, July 2013

[8] Sigal Berman, Member, IEEE, and Helman Stern, Member, IEEE," Sensors for Gesture Recognition Systems, IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 42, no. 3, may 2012

[9] Hwan Heo, Eui Chul Lee, Kang Ryoung Park, Chi Jung Kim, and Mincheol Whang," A Realistic Game System Using Multi-Modal User Interfaces", IEEE Transactions on Consumer Electronics, Vol. 56, No. 3, August 2010.

[10] Jashan M, Madhusoodhan H V, "Hand Gesture Recognition System Using Principal Component Analysis and its Applications", Electronics and Communication, SIT.

[11] Shinko Y. Cheng, Member, IEEE, and Mohan M. Trivedi, Fellow, IEEE "Vision-Based Infotainment User Determination by Hand Recognition for Driver Assistance", IEEE Transactions on Intelligent Transportation Systems, Vol. 11, No. 3, September 2010

[12] http://en.wikipedia.org/wiki/Gesture_recognition.

[13] Bobo Zeng, Guijin Wang, Xinggang Lin," A Hand Gesture Based Interactive Presentation System Utilizing Heterogeneous Cameras", TSINGHUA SCIENCE AND TECHNOLOGY ISSNll1007-0214ll15/18llpp329-336 Volume 17, Number 3, June 2012.

[14] Chun Zhu and Weihua Sheng," Wearable Sensor-Based Hand Gesture and Daily Activity

[15] Recognition for Robot-Assisted Living", IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, Vol. 41, No. 3, May 2011

[16] Lalit Gupta and Suwei Ma, "Gesture-Based Interaction and Communication: Automated Classification of Hand Gesture Contours", IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 31, no. 1, February 2001.

[17] H. K. ANASUYA DEVI, Fellow, "Thresholding: A Pixel-Level Image Processing Methodology Preprocessing Technique for an OCR System for the Brahmi Script", National Institute of Advanced Studies IISc Campus, Bangalore-12.

[18] http://www.evaluationengineering.com/articles/200608/blob-analysis-and-edge-detection-in-the-real-world.php

[19] Lamberti, L., & Camastra, F. (2011), "Real-time hand gesture recognition using a color glove", Springer-Verlag Berlin Heidelberg, ICIAP 2011, Part I, LNCS 6978, pp. 365-75.

[20] Bellarbi, A. ; Centre de Dev. des Technol. Av., Algiers, Algeria ; Benbelkacem, S. ; Zenati-Henda, N. ; Belhocine, M.," Hand gesture interaction using color-based method for tabletop interfaces", Intelligent Signal Processing , 2011 IEEE 7th International.

[21] Hitoshi Hongo*, Mitsunori Ohya and Kazuhiko Yamamoto," Focus of Attention for Face and Hand Gesture Recognition Using Multiple Cameras", IEEE Conf.

[22] http://www.silentlycrashing.net/p5/libs/jmyron/

[23] Gordana Lamb, "Improve Your UI Design Process with Object-Oriented Techniques", Visual Basic Developer magazine. 2001. quote: Table 1. Differences between the traditional application-oriented and object-oriented approaches to UI design."

[24] Monuri Hemantha, M.V.Srikanth," Simulation of Real Time Hand Gesture recognition for Physically Impaired", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2013.

[25] http://www.vision-systems.com/articles/print/volume-13/issue-9/technology-trends/image-capture/hand-gesture-recognition-system-targets-medical-applications.html

[26] Prof Kamal K Vyas, "Gesture Recognition and Control", International Journal on Recent and Innovation Trends in Computing and Communication ISSN 2321 – 8169, Vol.1.

[27] Sigal Berman, Member, IEEE, and Helman Stern, Member, IEEE, "Sensors for Gesture Recognition Systems", IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews, Vol. 42, No. 3, May 2012

[28] Bobo Zeng, Guijin Wang, Xinggang Lin," A Hand Gesture Based Interactive Presentation System Utilizing Heterogeneous Cameras", TSINGHUA SCIENCE AND TECHNOLOGY, 329-336,Volume 17, Number 3, June 2012

[29] Phillip Ian Wilson,Dr. John Fernandez, "Facial Feature Detection using HAAR Classifiers", JCSC 21, 4(April 2006)

[30] Archana S. Ghotkar & Gajanan K. Kharate ,"Hand Segmentation Techniques to Hand Gesture Recognition for Natural Human Computer Interaction", International Journal of Human Computer Interaction (IJHCI), Volume (3) : Issue (1) : 2012 15

[31] Q. Chen, N. Georganas, and E. Petriu, "Real-time vision-based hand gesture recognition using Haar-like features," in Proc. IEEE IMTC, 2007, pp. 1–6.

[32] S. Marcel, "Hand posture recognition in a body-face centered space," in Proc. Conf. Human Factors Comput. Syst. (CHI), 1999, pp. 302–303.

[33] W. Chung, X. Wu, and Y. Xu, "A real time hand gesture recognition based on Haar wavelet representation," in Proc. IEEE Int. Conf. Robot. Biomimetics, 2009, pp. 336–341.

[34] Y. Fang, K. Wang, J. Cheng, and H. Lu, "A real-time hand gesture recognition method," in Proc. IEEE Int. Conf. Multimedia Expo, 2007, pp. 995–998.

[35] L. Yun and Z. Peng, "An automatic hand gesture recognition system based on Viola-Jones method and SVMs," in Proc. 2nd Int. Workshop Comp.Sci. Eng., 2009, pp. 72–76.

[36] Y. Ren and C. Gu, "Real-time hand gesture recognition based on vision," in Proc. Edutainment, 2010, pp. 468–475.