Significance of Hand Gesture Recognition Systems in Vehicular Automation- A Survey

Jobin Francis
Dept. of ECE
Vimal Jyothi Eng College,
Chemperi, Kannur, India

Anoop B K
Dept. of ECE
Vimal Jyothi Eng College,
Chemperi, Kannur, India

ABSTRACT
Hand gestures are powerful human to human communication channel which convey a major part of information transfer in our everyday life. Hand gestures are the natural way of interactions when one person is communicating with one another and therefore hand movements can be treated as a non verbal form of communication. Hand gesture recognition is a process of understanding and classifying meaningful movements by the human hands. Nowadays vehicles launched from the industry offers an increasing number of infotainment systems as well as comfort functions that can be controlled by the driver. Though they are feature rich which demands more attention of the driver and degrade the driving performance and thereby reducing the safety. The gestural interaction is a promising means to cover the full range of driver's operational needs while minimizing the visual workload. Hand gesture recognition is of great importance for human computer interaction (HCI) because of its extensive applications in virtual reality and sign language recognition etc. Human hand is very smaller with very complex articulations comparing with the entire human body and therefore errors can be easily affected. It is thus a very challenging problem to recognize hand gestures. This paper comprises of the existing methods in detecting and recognizing hand gestures and a detailed study on their performances, accuracy, convenience, operational range and design challenges etc.

Keywords
Gestural interaction, Hand gesture recognition, Vehicular automations, HCI (Human Computer Interaction), FEMD, gesture vocabulary.

1. INTRODUCTION
Driving is a physical as well as a mental skill and it is the controlled operation and movement of a vehicle. Nowadays driving a modern car is a complex task. A driver has to continuously observe the road ahead and behind. Also he or she has to monitor the vehicle's status, operate on numerous electronic and mechanical devices to control the vehicle. And the driver assistance applications such as proximity distance control, GPS and audio entertainment etc which also consumes more visual attention. Hand gesture detection systems are one of the hottest fields of research since it is of great significance in designing artificially intelligent human-computer interfaces (HCI) [1] for vehicular applications. Emerging technologies now make it possible to create gestural -interfaces for vehicles that make it easier for the driver to carry the tasks effectively.

Almost all the hand gesture detection systems mostly follow the mentioning steps in detecting and recognizing the gestures. Image acquisition, preprocessing and hand gesture recognition. [8] Image acquisition involves fetching the movements of the hands by any image acquisition hardware and the captured gestures are then moved to the second pre-processing segment. Pre-processing methods [3, 8] are mainly based on the combinations of several processing operations. Pre-processing is a process that receives the data obtained from image acquisition steps and then prepares the data for the following procedures. This is done in a particular way as that the data can be more easily and effectively processed by the next detection steps.

Different pre-processing steps involve filtering, edge detection, histogram equalization, thresholding, etc.[8] Pre processing is the step which determines the accuracy of the system. Image acquisition modules are different related to different detection systems. The modules can be an accelerometer sensor or a data glove or a in vision based approach a camera replaces the above mentioned modules. The data fed from the image acquisition section may be mostly error affected .Therefore pre-processing is the step which enhance the quality of the information from acquisition step and ultimately improves the accuracy of the detection systems.

Recognition step identifies the processed hand movements as a particular gesture and then determines what the operation should be executed for a particular hand gesture [2]. Recognition step involves different algorithms for individual detection systems. In this paper it is described some of the gesture detection systems and the related algorithms used, results obtained, challenges faced etc. Rest of the paper is organized as follows. Section 2 describes the motivation and related work and section 3 presents some of the hand gesture detection systems, their recognition methods, applications and design challenges etc. Section 4 discusses a performance evaluative study and a hand gesture vocabulary is proposed and section 5 concludes the paper. This paper actually makes a comparative analysis between the different existing hand gesture detection systems and finally on comparing these systems pointing out the features of Kinect sensor based vision gesture detection systems in detail in the conclusion section of the paper.

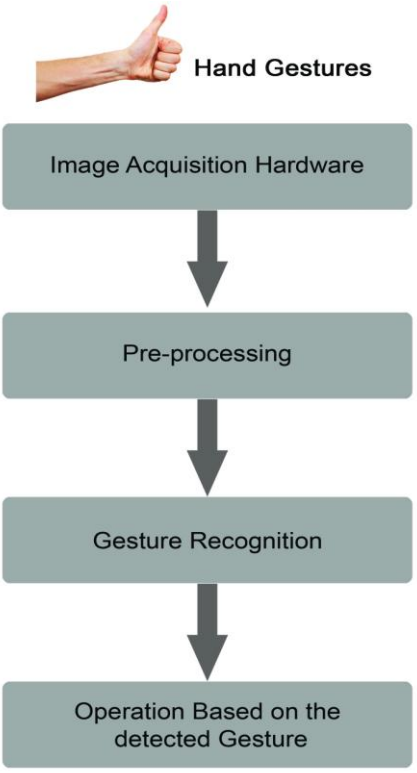


Figure 1
Hand Gesture Detection System

2. MOTIVATION AND PRIOR WORK

Many gesture detection systems have been applied successfully to different fields with some good results. The reason for choosing hand as the gesture detection input in vehicular automation is, the hand is more flexible than any of the body part and more no gestures that a hand can be generated than that can be generated by the head or face [1]. So gesture 'vocabulary' will be more if hand movements as the gesture. In many of the areas gestures replaces some of the conventional input devices as they are found more feasible than the conventional types in some particular areas or applications. This section mentions those people who proposed hand gesture detection systems, the fields that they applied and the types of acquisition module used etc.

Hand gesture detection started with the invention of glove based control interfaces.[9] Researchers realized that the gestures inspired by sign language used to offer simple commands for a computer interface [4] .Then the stream has been gradually evolved and widened with the invention of accelerometers , infrared cameras, depth sensors etc.[9] The first glove prototype developed in 1977 was Sayre Glove which used flexible tubes and light source at one end and a photo cell at the other.[9] Bending of the finger resulted in decreasing the amount of light. The amount of fingers bent can be determined by the measured voltage in the photo diode used. In 1983 Gary Grimes developed a digital entry data glove[9] the glove was consisting different sensors especially touch sensors to determine if the thumb was touched by another part of the head or finger.

In 1987 a commercially available data glove was introduced, equipped with 15 or more sensors with different gesture tracking abilities. Power Glove was developed in 1989 which was commercialized by Mattel Intellivision as control device for the video game console. The P5 glove an updated version of the power glove in 2002 provides users intuitive interaction with 3D and virtual environments. Robert Wang and Papovic [10] proposed a colored glove based hand tracking system with the custom pattern imprinted on it. The proposed system was in expensive and which simplified the pose estimation problem. Feng Wang and Chong Wah [8] proposed a real time gesture detection system for lecture videos. Daeho Lee and Park [2] presented a universal remote control system based on computer vision. The motion and skin color of the hands are utilized to detect the waving hands and requesting control commands. In the proposed system the remote control mode is set by a hand pointing toward an appliance to be controlled in. Then the hand region is tracked to recognize unique hand gestures. The three dimensional coordinate values are then computed to detect the hand gestures. Doring and Kern [12] proposed a gestural interaction system mounted on the steering wheel which will be treated as an additional interaction surface. Since the button displays are very close to the drivers hand they demand the improved safety of the driver. Ahamed foudzi and Kuzman [4] presented a real time hand gesture system for mobile robot control. Through that paper they proposed a real time vision based approach for mobile robot control by hand gestures in front of a static camera. It is done by a posing a hand gesture in the vision range of the robot and extracts out the information then which actuates the robot motors.

In skin color based edge detection algorithms major issue is the inaccurate hand segmentation and with the development of depth sensors which lead to correct the problems related with hand segmentation .With the invention of RGB–D based cameras [2,7] greater accuracy is then achieved in gesture detection systems. Guo presented a skin color detection and foreground detection model to track the dynamic or static hand under changing backgrounds. Paul presented a comparative study of depth image [1,2] based hand extraction and RGB image based hand extraction. Maisto and Panella [7] proposed an algorithm for the identification of finger tips using an RGB- D [3] camera. The results show a really good identification rate and a faster execution speed and the system can have the capability of detecting even the little movements of the hand.

3. GESTURE DETECTION SYSTEM

3.1 Accelerometer based approach

The proliferation in the field of microelectronics has inspired research in the field of accelerometer based gesture detection systems. This section deals with an accelerometer based gesture recognition system works on 18 gestures over 3700 traces. This is a three axis accelerometer system which employs dynamic time warping (DTW) and affinity propagation (AP) [11] for creating the exemplars. A general over view of the system is as follows.

The core of the system is dynamic time warping (DTW). Dynamic time warping is an algorithm for measuring the similarity between two temporal sequences which may vary in time or speed and calculates the optimal match between the two given sequences. In this system the acceleration of the hand is used as the data to represent .It is measured at different times using a single three axis accelerometer. Therefore the value of each gesture is a three column matrix

representing the acceleration of the hand in x, y and z directions. Temporal variation is one problem usually occurring that hand gesture differs from one person to another and even the same person cannot perform the same gesture once again. Therefore there is a need of gesture traces can either be compressed or stretched depending on the user and the speed of the hand movement. Here the system consists of two stages a training stage and a testing stage .A diagram of the system is shown in figure 2.

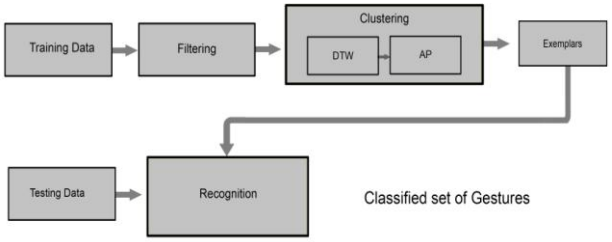


Figure 2. Accelerometer Based Gesture Detection System
DTW- Dynamic Time Warping AP- Affinity Propagation

The training stage consists of a filtering stage and a clustering process. The moving average filter is applied to the acquired data to remove any noise that might have been accumulated at the time of internal sampling, accelerometer calibrations, or sensitivity or hand shaking while the data acquisition etc.

In clustering process which uses the dynamic time warping (DTW) is used after done the smoothening process by the moving average filter. Dynamic time warping is the process to compare the measure of similarity between vectors of unequal length. The data from DTW [11] is given for affinity propagation which decomposes the training data into multiple clusters. Clustering is done in a particular way as members of the same cluster should have the same characteristics. Each cluster is represented by one of its members called an exemplar. Exemplar works as a reference model and the exemplars will then be given to the second testing process so that a set of exemplars makes a cluster of gesture traces. Recognition is carried out in two steps. First a set of exemplars which are very close to the observed data are selected and in the second step the best match among the members of the clusters form the first section is selected. The affinity propagation (AP)[11] algorithm is used and found effective as it simultaneously considers all the gesture data are exemplars and exchanges real valued messages among data points until a good set of exemplars and clusters emerges out. In the testing phase the unknown gesture trace is compared with the exemplars generated .The system is tested over 18 gestures which contains 3700 traces[11] collected from different subjects, and the system achieved almost perfect recognition for user dependant recognition ,mixed user and user independent recognition etc.

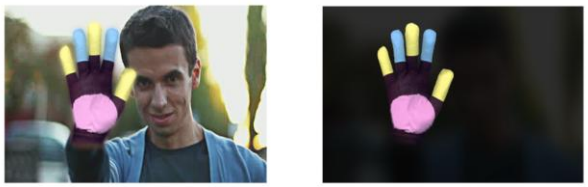
3.2 Glove based approach

Gesture detection systems are actually evolved with data glove based experiment. In section 2 while mentioning the related work, an overview of the evolution of data glove based detection systems has been given. Depending on the data glove used, the detection methods mainly are of two types. One is using the active data glove and second is using the passive data glove [9]. In active data glove based approach the glove consists of a variety of sensors can be an accelerometer sensor mounted and the communication is done by wired or -wireless with the parent device. The concept of accelerometer based approach that already discussed in section 3 under the main heading .Then in the passive data glove based approach the glove uses only markers or colors to distinguish the hand parts and which does not have onboard sensors on the glove. The gestures are fetched by an external camera. In this section it is dealt with the colored glove based gesture detection system[4]. The system is formed by three modules a camera to capture the gestures, a feature extractor, and a classifier which performs vector quantization.

In glove based approach the major part is the color glove on which different colors are specified to distinguish the different parts of the hands as palm, fingers etc. The palm has been given the color Magenta and colors Cyan and Yellow has been given to the fingers alternatively. Rest of the glove a black color is given as in the figure.

3.2.1 Segmentation process

The segmentation module that receives the RGB color frame acquired by the camera and performs the segmentation process to identify the hand image. In first step the RGB color frame is converted to Hue-saturation- Intensity color space[4]. It is followed by the thresholding process. The pixels of the image are divided into seven categories, Cyan Pixels (C) , Probable Cyan pixels (PC), Yellow pixels (Y), Probable Yellow pixels (PY), Magenta pixels (M), Probable Magenta pixels(PM), and black pixels(B).Thresholding is done by only pixels belonging to PC,PY,PM categories are considered and those categories are upgraded respectively to C,Y and M .The remaining pixels are degraded to black pixels and finally only four categories C ,Y ,M and B remains. Then the intermediate frame is as shown in figure 3 (b).



(a) (b)
Figure 3. (a) Original Image (b) Image after the segmentaion process

3.2.2 Feature extraction

The feature extraction step [4, 6] has the following steps. First the palm region is individuated and centroid of the palm region is computed and the major axis with respect to the palm region is drawn. Then the five centroids of the Yellow and Cyan region are then computed. The angles θi (i=1 to 5) between the mean axis and connecting centroids are computed

Thus the hand image is represented by nine normalized numerical features which gives five distances di (i=1 to 5) and four angles βi and the values are shown in figure 4 [4]. Each distance measures the Euclidian distance between the centroids of the palm and the respective finger. The angles between the fingers are computed and finally all the distance and angle values are normalized. Learning Vector quantization (LVQ) [4] is the step after the feature extraction

LVQ is a version of vector quantization which generates code vectors to generate decision boundaries.

Learning Vector Quantization methods are generally used in pattern classification problems. LVQ establishes a number of

-code book vectors and approximate the input values by some quantized values. In the data space the reference values of the gestures; the nine dimensional vectors are present .If they are agree then the system considers the input data as a gesture. LVQ algorithms are very easy to be implemented for multi class classification problems and complexity of algorithm can be adjusted. LVQ 1, LVQ 2, LVQ 3 [4] etc are some versions of this algorithms.

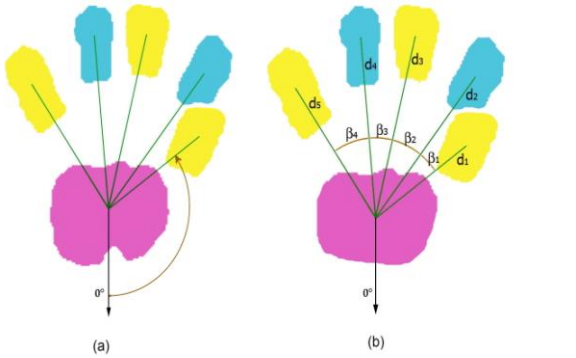


Figure 4. (a) Feature extracted image (b) Nine dimensional vector representation of the feature extracted image

3.2.3 Vision based approach using kinect sensor

The recent development of the depth cameras [2, 3] leads to new opportunities in hand gesture detection systems. The inexpensive depth cameras the Kinect sensors provide a valid base for gesture recognition. The input acquisition device Kinect Sensor is composed by an RGB camera[3] with 8 bit VGA resolution of 640× 480 pixels at 30 frames per second using a Bayer color filter.[3] The depth sensor uses an infrared laser projector combined with a monochrome sensor which captures the data under any lighting conditions. Microphones are also attached to execute vocal commands a tilt motor is also present so that the module can change position both in vertical as well as in horizontal directions. In Kinect sensor [1] approach both the color image and its corresponding depth map are taken. With the help of depth cue users hand can be detected. Depth perception is the ability to perceive the three dimensional view and the distance of an object. Usually the two fingers of the hand are indistinguishable as they are close to each other. Therefore recognizing such a noisy hand counters is overcome by a distance metric FEMD (Finger's Earth Movers Distance) [1] technique to measure the dissimilarity between the hand shapes. The different module in the Kinect sensor based detection system has given in the figure 5.

The Kinect sensor takes both the color image and the depth map at 640×480 resolution.[3] To segment the hand shape located the hand position by Kinect windows SDK hand tracking function. Then the hand region can be obtained by thresholding at a certain depth interval. The detection of the hand can be carried out very easily if the user wears black string on wrist. [1]Thus after detecting the black color pixels a line can be drawn to locate the black string. After detecting the hand shape represents those as a time series representation used for the classification and clustering of shapes. In time series representation the horizontal axis represents the angle between each contour vertex and the vertical axis denotes the Euclidian distance between the contour vertices and the centre point. The time series curve captures the topological properties of the hand. Time series curve shape representation

-is the successful method used for the classification and clustering of shapes.

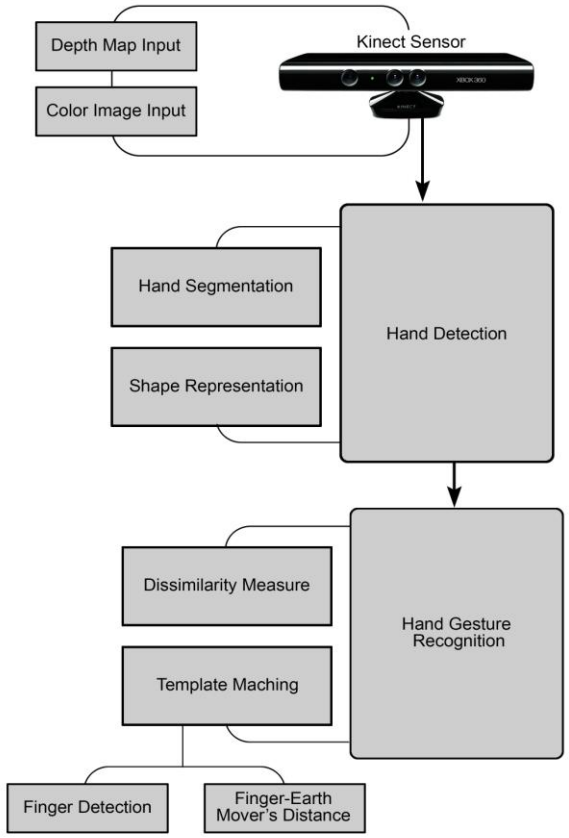


Figure 5. Kinect Sensor based gesture detection system

In gesture recognition step the system uses the method template matching. In template matching the input hand is recognized as the class which has the minimum dissimilarity distance.

c = argminc FEMD (H, Tc)

H is the input hand and Tc is the template of class c; FEMD (H, Tc) denotes the proposed Finger Earth Mover's Distance between the input hand and each template.EMD [1] is a measure of the distance between two probability distributions. The detection of the fingers is carried out by representing the hand shape as a signature with each finger as a cluster. Near – Convex Decomposition and thresholding decomposition [1] are two finger detection methods to obtain the finger parts from the hand shapes. Since the time-series curve gives the hand's topological information each finger corresponds to a peak in the curve. Kinect sensor based approach shows robustness to distortions and hand variations in orientation, scale and articulations since the FEMD metric [1] is insensitive to such distortions and variations. Since hand shape is detected using the depth information this system durable to cluttered backgrounds. Even if the hand image is cluttered by the other body part where the skin color is similar the detection process can be done effectively. Here in the figure 6 the first figure represents the hand image and the second one represents more detected version of the hand since it is detected with the help of black string tied over the hand. Image (c) represents its contour representation .the graphical representation actually shows the relative distance of each fingers and finger measurements. For the different hand

-movements the contour representation and the time series graph also will be differed.

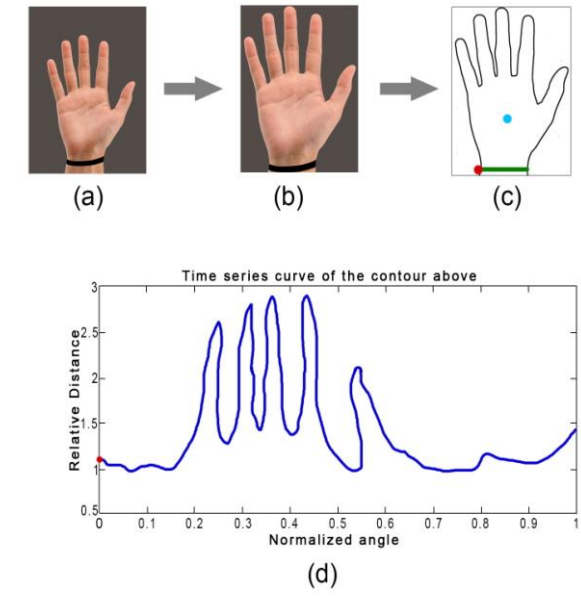


Figure 6. (a) Hand Image
(b) Detected hand with a black string
(c) Contour representation of the hand
(d) Time series representation of the hand image

In Kinect based approach even the depth map may be low resolution, the FEMD (Finger Earth Mover's Distance) based approach [1] discussed in the previous section the above problem can be rectified effectively and based on this method fingers can be detected and this could help to produce a versatile set of gesture input for a vehicle automation system. Hence Kinect based gesture detection system is the best method for vehicular automation and the gesture 'vocabulary' proposed in this section that can be used as the different gesture input in automating the features in vehicles.

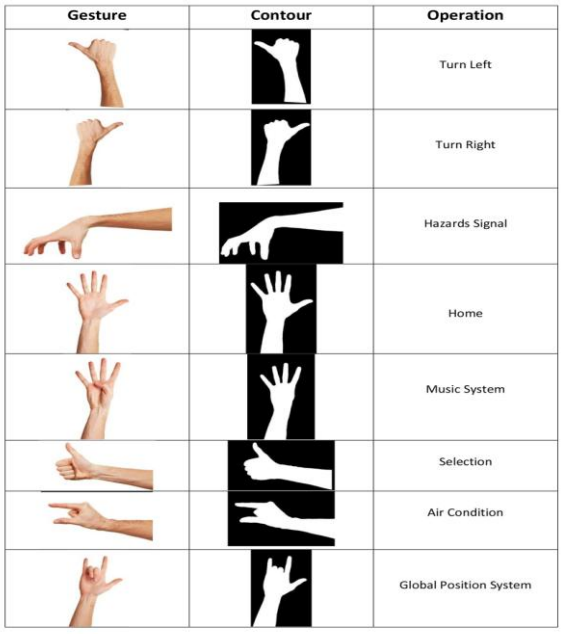


Figure 7. Gesture Vocabulary

4. PERFOMANCE EVALUATION

Hand gesture detection systems confronts with many challenges. Variations in the illumination such as lighting condition affects at the time of extracting the hand skin region. If the hand region is rotated in any direction or any other objects in scene in mingled with the hand to be detected may disrupt the total performance of the system. This section carries a performance evaluative study on the various gesture detection systems discussed in the previous sections.

Even though the accelerometer based and the glove based approach produced good results in some scenarios successfully, the Kinect sensor [1, 3] based approach is found effective in implanting in vehicular automation field. Since in the first two methods mentioned here the driver or the user has to wear a glove or a bundle of sensors in hand, which may seriously affect one's driving. Therefore vision based hands free gesture detection systems are a better option in vehicular application and hence the Kinect sensor based method is the best suit. Kinect sensor shows a good robustness cluttered backgrounds and also to distortions and hand variations while acquiring the input image. The glove based method may be user dependent as the glove once designed may be inappropriate for another user that hand size differs from user to user. Also it take s time to put on or take off which also may interfere with driving and reduce the driving performance. In accelerometer based approach the user must attach more no of sensors to hand or arm therefore such system offers very limited movements of the hands. Also accuracy can be a problem that the variations occurred in sensor output due to a temperature and earth's magnetic field.

The operational range is found as some meters in the case of wireless sensors of this type and the range is limited up to the cable length in wired based approach.

5. CONCLUSION

Hand gesture detection for real life applications is a very challenging because of its robustness accuracy and efficiency. This paper has been formulated an evaluative study on some gesture detection systems, their mode of operation, system description and some results etc. The paper actually aims at to reveal the significance of gestures in vehicular applications. The major contribution of this Kinect based approach is the distance metric based on the part-based representation. The future work can be focused on implementing a hand gesture detection system in vehicles that can access features using the bare hand while driving with the intention of reducing the visual demand of the driver in operating the vehicle applications and thereby increase the driver's safety. The tool which we plan to use is open CV with the aid of Raspberry- pi micro controller.

6. REFERENCES

[1] Zhou Ren, Jingjing Meng and Zhengyou Zhang, "Robust Part-Based Hand Gesture RecognitionUsing Kinect Sensor", IEEE Transactions on multimedia, vol. 15, no. 5, august 2013

[2] Yuan Yao and Yun Fu, "Contour Model based Hand-Gesture Recognition Using Kinect Sensor", 10.1109/TCSVT.2014.2302538, IEEE Transactions on Circuits and Systems for Video Technology.

[3] Juan P. Wachs, Helman Stern and Yael Edan, "Cluster Labeling and Parameter Estimation for the Automated Setup of a Hand-Gesture Recognition System", IEEE

Transactions on Systems, Man, and Cybernetics—part a: Systems and Humans, vol. 35, no. 6, november 2005

[4] Luigi Lamberti and Francesco Camastra, "Real-Time Hand Gesture Recognition using a Color Glove" ICIAP 2011 Part 1 LNCS 6978 pp-365 2011

[5] Harshith.C, Karthik.R.Shastry, Manoj Ravindran, M.V.V.N.S Srikanth and Naveen Lakshmikhanth, "Survey on various gesture recognition techniques for interfacing machines based on ambient intelligence", International Journal of Computer Science & Engineering Survey (IJCSES), Vol.1, No.2, November 2010

[6] Daeho Lee and Youngtae Park, "Vision-Based Remote Control System by Motion Detection and Open Finger Counting" IEEE Transactions on Consumer Electronics, Vol. 55, No. 4, November 2009

[7] Marco Maisto, Massimo Panella, Luca Liparulo, and Andrea Proietti, "An Accurate Algorithm for the Identification of Fingertips Using an RGB-D Camera", IEEE Journal on emerging and selected topics in circuits and systems, vol. 3, no. 2, June 2013

[8] Regina Lionnie, Ivanna K. Timotius & Iwan Setyawan, "Performance Comparison of Several Pre-Processing Methods in a Hand Gesture Recognition System based on Nearest Neighbor for Different Background Conditions" ITB J. ICT, Vol. 6, No. 3, 2012, 183-194

[9] P. Premaratne, Human Computer Interaction Using Hand Gestures, Cognitive Science and Technology, DOI 10.1007/978-981-4585-69-9_2, © Springer Science+ Business Media Singapore 2014

[10] Robert Y. Wang and Jovan Popovic, "Real-Time Hand-Tracking with a Color Glove",

[11] Ahmad Akl Chen Feng and Shahrokh Valaee ," A Novel Accelerometer-Based Gesture Recognition System" IEEE Transactions on Signal Processing, vol. 59, no. 12, december 2011

[12] Tanja Döring, Dagmar Kern, Paul Marshall and Max Pfeiffer, Johannes Schöning, Volker Gruhn and Albrecht Schmid, "Gestural Interaction on the Steering Wheel – Reducing the Visual Demand",

[13] Siddharth S. Rautaray, Anupam Agrawal, "Real Time Multiple Hand Gesture Recognition System for Human Computer Interaction", I.J. Intelligent Systems and International Journal of Computer Applications, 2012, 5, 56-64 Published Online May 2012 in MECS DOI:10.5815/ijisa.2012.05.08 http://www.mecs-press.org

[14] Jiayang Liu, Zhen Wang and Lin Zhong, "uWave: Accelerometer-based Personalized Gesture Recognition and Its Applications" Pervasive and Mobile Computing 5 (2009) 657-675 in www.elsevier.com/locate/pmc

[15] Feng Wang, Chong-Wah Ngo and Ting-Chuen Pong, "Simulating a Smartboard by Real-Time Gesture Detection in Lecture Videos" IEEE Transactions on Multimedia, vol. 10, no. 5, august 2008

[16] Andrew D. Wilson and Aaron F. Bobick, Member, "Parametric Hidden Markov Models for Gesture Recognition" IEEE transactions on pattern analysis and machine intelligence, vol. 21, no. 9, September 1999

[17] Andreas Riener and Johannes Kepler University of Linz, "Gestural Interaction in Vehicular Applications",

[18] Ahmad Athif Mohd Faudzi, Muaammar Hadi Kuzman Ali, M. Asyraf Azman and Zool Hilmi, "Real-time Hand Gestures System for Mobile Robots Control", International Symposium on Robotics and Intelligent Sensors , IRIS 2012