

PAPER • OPEN ACCESS

An Overview of Hand Gestures Recognition System Techniques

To cite this article: Farah Farhana Mod Ma'asum *et al* 2015 *IOP Conf. Ser.: Mater. Sci. Eng.* **99** 012012

View the [article online](#) for updates and enhancements.

An Overview of Hand Gestures Recognition System Techniques

Farah Farhana Mod Ma'asum¹, Suhana Sulaiman, Azilah Saparon
Center for Electronics Engineering Studies (CEES)
Faculty of Electrical Engineering, Universiti Teknologi Mara Shah Alam,
40450 Shah Alam, Selangor, Malaysia

E-mail: farah.farhana88@yahoo.com

Abstract. Hand gesture recognition system has evolved tremendously in the recent few years because of its ability to interact with machine efficiently. Mankind tries to incorporate human gestures into modern technology by searching and finding a replacement of multi touch technology which does not require any touching movement on screen. This paper presents an overview on several methods to realize hand gesture recognition by using three main modules: camera and segmentation module, detection module and feature extraction module. There are many methods which can be used to get the respective results depending on its advantages. Summary of previous research and results of hand gesture methods as well as comparison between gesture recognition are also given in this paper.

1. Introduction

Massive improvement in computer development can be clearly seen involving both interaction between human and computer itself. Human interaction with computer has becoming a core component in our daily life as we interact with it throughout the day. Technology nowadays incorporate flat panel displays as an output device of images being transmitted electronically. For an instance, most cars are equipped with computerized systems for navigation and entertainment which needs to be operated and displayed by using a screen panel. However, controlling a computer by gesturing in the air will be everyone wishes as this will ease down our tasks. Hand gesture has been visualized as the new evolvement of human computer interaction which will replace the functions and usage of touch screen displays.

Human computer interaction (HCI) is defined as the relation between the human and computer (machine), represented with the emerging of computer itself [1]. Vital endeavour of hand gesture recognition system is to create a natural interaction between human and computer when controlling and conveying meaningful information. There are two main characteristics should be considered when designing a HCI system as mention in [1]: functionality and usability.

System functionality referred to a set of actions or services that are provided to its users [1], while usability of a system is defined by the level and scope of the system in which the system can be used efficiently in order to obtain certain specific user purposes [2]. By having an equal balance between both functionality and usability of a system, an effectiveness and powerful performance system can be achieved. However in order to create a useful hand gesture recognition system,



differences between hand postures and hand gestures need to be distinguished first. Hand posture is a static hand configuration which is represented by a single image without any involvement of movements [3][4], while hand gesture is defined as a dynamic movement consisting of sequences of hand postures over a short span of time [4]. For example, making thumbs up and holding it in a certain position is a hand posture while, waving good-bye is an example of hand gesture.

Gestures by using sign language are usually used among deaf community for communication between human and computer [5]. Hand gestures may represent compressed information [6] which can be decoded between both sender and receiver in order to obtain and understand the gesture meaning. This paper intends to provide an overview of hand gesture recognition system which addresses different approaches that can be used to acquire information of hand gestures recognition, which are segmentation, features extraction, and recognition.

Most of the studies have defined specific gestures for a specific task based on its own purpose of studies. In research done by S. Billingham *et. al* [7], natural gestures for common web browsing tasks are tested by examining differences between postures, screen sizes and user's educational background. S. Billingham claims that there are quite distinctive differences between the psychology and engineering participants. The psychology and engineering groups has different ways of swiping touch screen panel. Psychology group preferred to swipe from right to left while the engineering group preferred the opposite directions.

2. System components and framework

In order to analyse and process hand gesture recognition, it needs to undergo three main steps which are: camera module, detection module and feature extraction module, as shown in Figure 1. Camera module acquires images from input source while detection module is where the quality of the threshold images is determined. Feature extraction module on the other hand acts as extraction of user's gesture according to its purposes. Comparison between each researcher methods is further explained in Section 2.1.

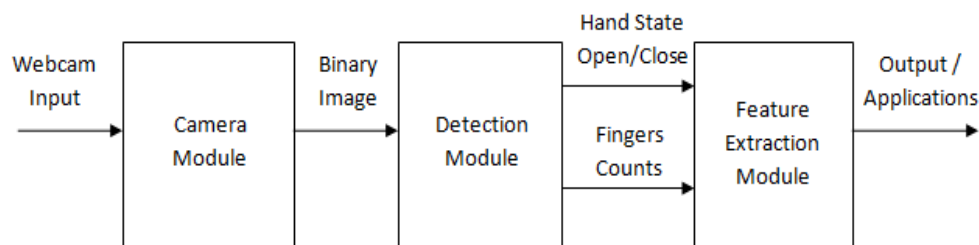


Figure 1. Framework of hand gesture recognition system.

2.1. Camera & Segmentation Module

In this camera module, image of hand can be retrieved from various sources such as single camera (web camera), depth camera and infrared camera with sensors. The image of hand is then need to be segmented so that it could divide input image into required regions. It dealt with separating the user's hand with background in order to capture the gestures without noises.

However, both hand postures (static) and hand gestures (dynamic) do required different segmentation process depending on the type of gestures. If it is dynamic gesture, it needs to undergo two steps before it is being segmented. Firstly, the hand gesture needs to be located and tracked using a bounding box, then the video of gestures will have to be divided into frames and each frame have to be processed alone [8].

A perfect segmentation is difficult to obtain due to overlapping of skin colours with complex background. There are various researches which has been done using different type of colour space technique in segmentation process such as HSV colour model, YCbCr colour space [9] and RGB colour model [10]. However, due to sensitivity of lighting changes in colour spaces, many researches tend to eliminate the luminance component [11].

The most important step for hand segmentation is thresholding which it is used to separate hand from the background. Thresholding can be used as an extraction method from the background by assigning certain values for each pixel as a threshold value[10]. Thresholding will produce a binary image with all pixels of 1 representing the hand. This occurs when pixel with intensity less than the threshold value is set to 0, while any pixel which has the intensity more than the threshold value will be set to 1. Figure 2 shows the results of hand segmentation by using RGB colour space with different threshold values, 70 and 20, by A. Dhawan *et. al* [10] .



Figure 2. Segmentation using RGB color space with different threshold value (left: original image, center: threshold value of 70, right: threshold value of 20) by A. Dhawan [9]

Figure 3 shows the results when H.S. Yeo *et. al* [9] applied YCbCr colour model with different wide and narrow YCbCr threshold value. It can be seen clearly that background with similar colours as skin is not extracted when using a narrow threshold.



Figure 3. Using wide YCbCr threshold (left: original image, right: after skin extraction), by H.S. Yeo [8]

2.2. Detection Module

Detection module is a module where is responsible for detecting and tracking hand and fingers. Fingers and hand locations, hand close/open state and fingers direction need to be extracted in order to gain more information about the gestures.

2.3. Feature Extraction Module

Features vector of the segmented image can be extracted in many ways according to particular application. Some methods used the shape of the hand such as hand contour which is then being calculated for convex hull and convexity defects [10], while M.M. Hassan *et. al* [5] prefer to divide the segmented image into different blocks representing its brightness measurement of the image. Center of gravity (COG) [12] of segmented hand and the distance from COG to the farthest point of the fingers, is another method which can be used as an feature extraction.

3. Literature Review on Gesture Recognition Systems

M.M. Hassan *et. al* [5] applied scaled normalization for gesture recognition based on brightness factor matching. Initially, the input gesture image is divided into 25x25 blocks (each of 5x5 block size), which is then colour segmented by HSV colour model and thresholding technique. Each segmented

images are normalised by using centre of mass technique as shown in Figure 5. In Figure 5, the coordinates of red line are shifted to match the centroid of the hand object at the origin X and Y axis. After being normalized according to its respective normalization technique, these images are then being analyzed by using local brightness value mathematical equation[5]. Comparison between database features and tested gesture features are then being calculated using equation (1) for the percentage between feature image and database, as in [5],[13]. The database consists of 6 different gestures, while 10 samples per gesture are used (5 samples for training and 5 samples for testing). And from his outcome, it has shown that the recognition rate for center mass normalization, as in [5] achieved better results compared to classical normalization technique, as in [13]. This is because of its ability to extract more data from hand information itself. However, time limitation is the disadvantage of this research since it consume more time in interpreting and comparing each gestures and its database.

$$Recognition (\%) = \frac{\text{number of matched features}}{\text{total number of blocks}} \times 100\% \quad (1)$$

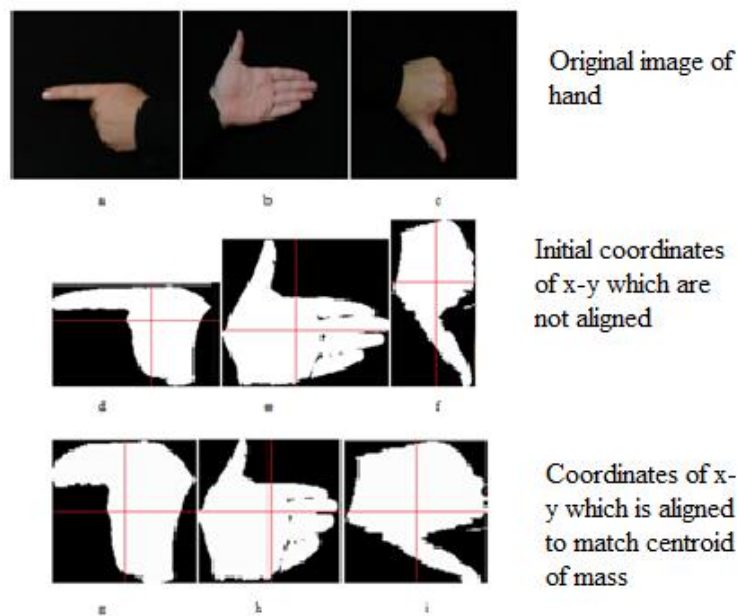


Figure 4. Applying scaling normalization on the input image (a,b,c: Original Image. d,e,f: Image after being segmented. g,e,f: image after scaling normalization process) [5]

G. Thomas *et. al* [14] on the other hand had compared three different types of methods which can be used to recognize hand gestures. These three methods are pixel-by-pixel comparison, edges method and orientation histogram. Pixel-by-pixel comparison is a method where each and every frame is being compared with the image in the database. Edges method [14] on the other hand is used to find out which portion of the image produce the highest gradient value. In order to determine edges of the hand image, two filters named X-Y filter is used – see Figure 6. These two filters represent different directions respectively, for the x direction, $X = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix}$ while for the y direction, $Y = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$.

Orientation histogram method is actually a method where it combines both edges method and pixel-to-pixel comparison method. Each pixels of the image (x,y) has its own gradient. Gradient direction and gradient magnitude are given by $\arctan(dx,dy)$ and $\sqrt{dx^2 + dy^2}$ respectively, as in [14]. If the gradient magnitude is greater than threshold, the gradient direction is detected and the frequency is incremented. This histogram is then saved as a training pattern database. As the last steps, all three

methods (pixel-by-pixel comparison, edges method and orientation histogram) used Euclidean Distance to determine the differences between image and its database. And from this finding, it shows that orientation histogram produced the highest percentage of recognition since it is able to recognize gesture fast and robust. However, it does have disadvantage of being rotation dependent.

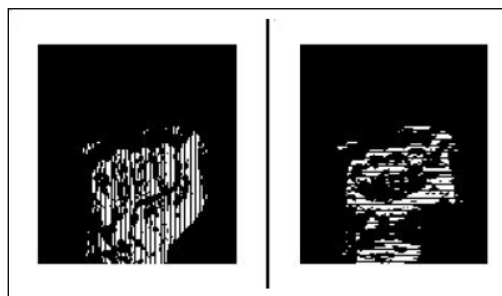


Figure 5. Edge detection using X-Y filter [13]

H.S. Yeo *et. al* [9] and A. Dhawan *et. al* [10] on the other hand implies quite similar technique of feature extraction in their research. Both researchers employed convex hull and convexity defects method to determine both hand and fingertips contours. However, they used difference type of input sources in their researches. A. Dhawan *et. al* [10] only focuses on single camera as an input source, while H.S. Yeo *et. al* [9] compares the results of image recognition using single camera and depth camera. In H.S. Yeo *et.al* [9] studies, he has proved that user can interact with computer without using any physical controllers. Another findings by H.S. Yeo *et. al* [9], single camera does not works well under low light condition. In order to overcome this issue, a large range of YCbCr threshold value has to be used. Nevertheless, this limitation does not occur when using Kinect depth camera due to its stability to keep the hand contours, by applying a fixed depth threshold.

Table 1 shows summary of some hand gesture recognition system: comparison between methods in hand gesture recognition methods are used, while Table 2 tabulates the summary of recognition system which used other tracking devices such as gloves etc.

Table 1. Summary of Hand Gesture Recognition Systems Using Different Types of Camera

Research	Method	Detection	Feature Extraction	Recognition (%)
M.M. Hassan, P.K. Mishra [13]	Vision-based using single camera	HSV colour space/ Threshold method	Classical normalization: Divide the hand image into blocks of intensity features, which is then being extract by using edge information technique	Classical normalization 91%
M.M. Hassan & P. K. Mishra [5]	Vision-based using single camera	HSV colour space/ Threshold method	Comparison between classical normalization technique and block scaling using center of mass normalization technique	Classical normalization 83.3% Block scaling normalization using center of mass 96.6%
Ginu Thomas [14]	Vision-based using single camera	RGB colour space/ Otsu threshold method	Image captured are being tested in few methods: a) Pixel by pixel comparison b) Edges method	Pixel by pixel method 86%, Edges method 92% Orientation

			c) Orientation histogram	histogram 93.6%
H.S. Yeo, B.G. Lee, H. Lim [9]	Vision-based comparison using single camera and depth camera	YCrCb colour space/ Threshold method	Fingertips recognition by using convexity defects method and convex hull method around hand contours	Depth camera produced better result compare to single camera
A.Dhawan, V. Hanrao [10]	Vision-based using single camera	RGB colour space/ Threshold method		

Table 2. Summary of Hand Gesture Recognition Systems Using Tracking Devices

Research	Method	Detection	Feature Extraction	Recognition (%)
Y. Yao, Y. Fu [15]	14- patches colour glove	RGB-D using Kinect sensor	Database indexing technique is used by collecting training samples. Similarity between training samples and gesture are tested	51.87%
J. Nagi, A. Giusti, L. Gambardella, G. A. Caro [16]	Coloured Glove	HSV colour space	Support Vector Machine (SVM) is trained in a cascaded multi-binary-class configuration, where the spatial gestures are effectively learned and recognized by a swarm of UAVs.	Spatial gesture recognition is robust and scales well with swarm sizes of up to 20 robots
Kılıboz, Nurettin Çağrı Gündükbay, Uğur [17]	Six-degree-of-freedom (DOF) magnetic motion tracking device	Collected motion data is converted into relative position data in gradient form (x-y axis)	Finite State Machine (FSM) is used as a gesture recognizer method. Needleman-Wunsch sequence matching algorithm is applied in order to produce similarity scores between two sequences.	Recognition rate: 73%. The outcome can be improved by using a more powerful tracking approach which has better sensor capability

4. Conclusion

In this paper, three main steps of hand gesture which are camera module, detection module and feature extraction module are discussed. A robust hand and finger gesture tracking recognition system is viable to be used in human computer interaction system. The purpose of this paper is to review other research studies on different types of detection module and feature extraction methods, as tabulated in Table 1 and Table 2. Each method has its own advantages and disadvantages to be considered. The selection of specific recognition extraction depends on the application needed. It can be concluded that depth camera do produced the best result in hand gesture recognition, and this is because it can be easily calibrated to suit any skin colour and it can work against almost any background.

5. Proposed Future Work

For future work, we would like to focus on implementing a robust marker-less hand gesture recognition system using a low-cost hardware. A sliding tracker will be used for the user to determine the threshold value so that it can easily suits any user's skin color by controlling RGB color

segmentation. Convex hull and convexity defects method on the other hand will be used as extraction methods to get the hand contours. Various gestures types will be tested in the future work such as pointing, clicking, swiping and scrolling.

5. Acknowledgement

Thank you to Kementerian Pelajaran Malaysia and UiTM for Tenaga Pengajar Muda (TPM) scholarship for funding this research.

References

- [1] F. Karray, M. Alemzadeh, J. a Saleh, and M. N. Arab, "Human-Computer Interaction: Overview on State of the Art," *Int. J. Smart Sens. Intell. Syst.*, vol. 1, no. 1, pp. 137–159, 2008.
- [2] J. Carey and M. Tremaine, "Integrating Human-Computer Interaction Development into the Systems Development Life Cycle: A Methodology," *Commun. Assoc. Inf. Syst.*, vol. 15, pp. 512–543, 2005.
- [3] G. Murthy and R. Jadon, "A review of vision based hand gestures recognition," *Int. J. Inf. Technol. Knowl. Manag.*, vol. 2, no. 2, pp. 405–410, 2009.
- [4] P. Garg, N. Aggarwal, and S. Sofat, "Vision based hand gesture recognition," *World Acad. Sci. Eng. Technol.*, pp. 972–977, 2009.
- [5] M. M. Hasan and P. K. Misra, "Brightness Factor Matching For Gesture Recognition System Using Scaled Normalization," *Int. J. Comput. Sci. Inf. Technol.*, vol. 3, no. 2, pp. 35–46, 2011.
- [6] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 37, no. 3, pp. 311–324, 2007.
- [7] S. S. Billinghamurst and K.-P. L. Vu, "Touch screen gestures for web browsing tasks," *Comput. Human Behav.*, vol. 53, pp. 71–81, 2015.
- [8] N. Ibraheem, R. Khan, and M. Hasan, "Comparative Study of Skin Color based Segmentation Techniques," *Int. J. Appl. Inf. Syst.*, vol. 5, no. 10, pp. 24–38, 2013.
- [9] H.-S. Yeo, B.-G. Lee, and H. Lim, "Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware," *Multimed. Tools Appl.*, May 2013.
- [10] A. Dhawan and V. Honrao, "Implementation of Hand Detection based Techniques for Human Computer Interaction," *Int. J. Comput. Appl.*, vol. 72, no. 17, pp. 6–13, 2013.
- [11] R. Zaman Khan and N. A. Ibraheem, "Hand Gesture Recognition : A Literature," *Int. J. Artif. Intell. Appl.*, vol. 3, no. 4, pp. 161–174, 2012.
- [12] A. A. M. Faudzi, M. H. K. Ali, M. A. Azman, and Z. H. Ismail, "Real-time hand gestures system for mobile robots control," *Int. Symp. Robot. Intell. Sensors*, vol. 41, no. Iris, pp. 798–804, 2012.
- [13] M. M. Hasan and Pramod K. Misra, "HSV Brightness Factor Matching for Gesture Recognition System," *Int. J. Image Process.*, vol. 4, no. 5, pp. 456–467, 2011.
- [14] G. Thomas, "A Review of Various Hand Gesture Recognition Techniques," *VSRD Int. J. Electr. Electron. Commun. Eng.*, vol. 1, no. 7, pp. 374–383, 2011.
- [15] Y. Yao and Y. Fu, "Contour Model-Based Hand-Gesture Recognition Using the Kinect Sensor," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 11, pp. 1935–1944, 2014.
- [16] J. Nagi, A. Giusti, L. M. Gambardella, and G. A. Di Caro, "Human-Swarm Interaction Using Spatial Gestures," *IEEE Int. Conf. Intell. Robot. Syst. (IROS 2014)*, no. Iros, pp. 3834–3841, 2014.
- [17] N. Ç. Kılıboz and U. Güdükbay, "A hand gesture recognition technique for human-computer interaction," *J. Vis. Commun. Image Represent.*, vol. 28, no. January, pp. 97–104, 2015.