An efficient hand gestures recognition system

Ahmed Kadem Hamed AlSaedi¹,² and Abbas H. Hassan AlAsadi³

¹ Faculty of Information Technology, University of Babylon, Iraq
² Faculty of Education, University of Misan, Iraq
³ Faculty of Computer Science and Information Technology, University of Basrah, Iraq
it.ahmedkadem@uobabylon.edu.iq
mr.ahmed@uomisan.edu.iq
abbas.hassin@uobasrah.edu.iq

Abstract. Talking about gestures make us return to the historical beginning of human communication because there is no language completely free of gestures. People cannot communicate without gestures. Any action or movement without gestures is free of real feelings and cannot express the thoughts. The purpose of any hand gesture recognition system is to recognize the hand gesture and used it to transfer a certain meaning or for computer control or/and a device. This paper introduced an efficient system to recognize hand gestures in real-time. Generally, the system is divided into five phases, first to image acquisition, second to pre-processing the image, third for detection and segmentation of the hand region, fourth to features extraction and fifth to count the numbers of fingers for gesture recognition. The system has been coded by Python language, PyAutoGUI library, OS Module of Python and the OpenCV library.

1. Introduction

Historically, the Electronic Visualization Lab was the first to create a data glove called Sayre Glove this was in 1977. Thirty-five years later, the researchers adopted the camera to interact with the computer. In fact, the camera is compared to the data glove and it is considered more direct and natural to achieve Human-Computer Interaction [1].

Recently, the interactive by gesture has become widely used and in the future may replace the mouse and/or keyboard by vision-based devices. The main feature of using hand gestures is to interact with the computer as an input unit. The gesture is defined as a form of nonverbal communication or non-vocal communication where the body's movement can convey certain messages. Gestures are originated from different parts of the human body, but the most common ones emerge from the hand or face.

Gesture provides a new form of interaction that reflects the experience of the user in the actual world. The interaction by the gesture is more natural and does not require any hindering or additional hardware.

There are two kinds of hand gestures, static and dynamic gestures. In [2] Liang introduced the best definition for static hand gestures (hand posture) and dynamic hand gesture as:
"Posture is a specific combination of hand position, orientation, and flexion observed at some time instance."
“Gesture is a sequence of postures connected by motion over a short time span.”
The good examples of static hand gesture are "OK" or "STOP" and "No", "Yes", "goodbye" for dynamic gestures.
Three approaches were used to obtain important information for any hand gesture recognition system. Data glove approaches, Vision-based approaches and Colored-Marker approaches as shown strictly in Figure (1).

In Vision-based approaches, the human motion obtained from one camera or more and devices based on vision can handle many properties for interpreting the gesture, for example, color and texture whereas the sensor has not this property [3] [4]. Although these approaches are simple, many challenges can appear for example, the lighting diversity, complex background, and present objects with skin-color similar to the hand (clutter), as well as the system, requires some criteria like time of recognition, speed, durability, and efficiency of computation [5] [6].

Data glove approaches use types of sensors to capture the position and motion of the hand. These approaches can compute in easy and accurate the coordinates of the locations of the palm’s fingers, and hand configurations [6] [7] [8]. The sensors do not achieve an easy connection with the computer because it needs to be the user connected physically with the computer and hinder the movement of the hand. These devices are also expensive and unsuitable to operate in an environment of virtual reality [8][9]. According to Moore's Law, the sensors will become smaller and cheaper over time. They will be prevalent in the future, we believe.

Colored- Markers approaches used marked gloves worn by the hand of a human and be colored to help in the process of hand tracking and to locate the fingers and palm. Marker gloves can form the shape of the hand by extract the geometric-features [10]. In [11] used a wool glove with three different colors to represent the palms and fingers. This approach considers simple and not expensive if compared with the sensor or Data Glove [11], but the natural interaction between humans and computers still not enough [9].

However, there are some challenges when we want to design a strong and real-time gesture recognition system. Challenges are related to complex hand structures that lead to difficulty in tracking and recognition. In addition, there are challenges related to the shape of the gesture, different lighting conditions, the real-time issue and the presence of noise in the background. These challenges are taken into account in this paper, by using the running average principle in the background subtraction technique to detect and extract the hand from the background and used the contour of hand as a feature.

(a) "Data-Glove approaches" [7]. (b) "Vision-Based approaches ". (c)"Colored Marker approaches"

Figure 1. Examples of input techniques for the hand gesture recognition system.

2. Related works

There are many systems designed to recognize the hand gesture, some of them will be mentioned. In [12], Amiraj and Vipul introduced a system to recognize the hand gestures for HCI. They used more than one approach for the preprocessing step in their system and used two methods to perform segmentation process one with a static background and another without the constraint of background. In the static background used constant value to a threshold, they used the Otsu thresholding algorithm, and when the thresholding is dynamic used the color at the real-time mode. In the free background constraint, used thresholding methods based on color feature and subtraction of the background model. For detecting the hand, they found the contour of it and then computing the convex hull and convexity
defect to find the number of fingers. They provided three approaches to interact with devices, tracking of the finger, the orientation of the hand, and counting of fingers.

Shwetha et al. [13] provided a review for many hand gesture recognition systems that used MATLAB language. They used a Canny edge algorithm to determine the edge of the hand and used the values of hue and saturation for Skin–Color detection. They concluded that the system gave better results when used Artificial Neural Networks (ANNs) and edge detection methods.

In [14], Nancy et al. introduced a system to hand gesture recognition by using the color-marker approach. The user wearing white cloth in his hand and place a red color marker on hand's fingertip. The gestures in this system used to point on the computer screen by detection the only finger with a red color marker. However, this system does not achieve direct contact with the devices because of the use of the color marker.

In [15], Tasnuva Ahmed introduced a hand gesture recognition system. The system based on neural networks and in real-time. The researcher divided the system into four steps, image capturing, pre-processing, features extraction, and recognition. The system succeeded in distinguishing hand gestures that taking from different angles or sizes or orientations, but there is a delay in the system due to the training phase of the Artificial Neural Network as well as delay in switching between the nodes.

Badgujar et.al [16] presented a recognition system for dynamic hand gestures using contour analysis. It is an efficient system for computer control, but it applies only to PowerPoint presentations.

Nagarajan et al. [17] introduced a system to recognize the gestures in real-time depending on the number of fingers from one to five. The system is divided into four phases; the first phase to capture the image in real-time by the camera. The second phase to segment the region of hand, using HSV color space and followed that performed operations of morphology. In the third phase, the contour of hand is detection by the convex hull approach. Finally, recognize the gesture according to the number of fingers. The pose orientation of hand considers the weakness of this system.

3. System architecture
The general structure for any system to recognize the hand gestures can be explained as shown in Figure (2).

![Figure 2. The general structure of the hand gestures recognition system.](image)

4. Proposed system architecture
The proposed system for hand gesture recognition consists of five phases. As shown in Figure (3), the block diagram of the system architecture.
The system received the hand gesture as input and execution a certain action that associated with this gesture as output. The algorithm for the system is shown below.

Start: Starting the camera
Step 1: Capture image from the camera.
Step 2: Extract the region of interest.
Step 3: Convert the RGB image to the grayscale image.
Step 4: Smooth the image by the Gaussian blur.
Step 5: Perform Background Subtraction.
Step 6: Threshold the image.
Step 7: Perform Erosion and Dilation from morphological operations.
Step 8: Extract the hand’s contour.
Step 9: Recognize the gesture by using two methods, Convex Hull and Convexity Defects.
Step 10: Execute the action that assigned to recognize gesture by computer.
Stop.

4.1. The image capturing
This phase used a webcam to acquire the image (frame by frame) and based on the only bare hand without a glove or colored marker that can be hinder the user.

4.2. Pre-processing
In this phase, in order to minimize the computation time, we took only the important area instead of the whole frame from the video stream, and this is called Region Of Interest (ROI). In the image processing prefers to convert the color images into grayscale images to increase processing and after complete the processing can restore the image to its original color space, therefore, we convert the region of interest into a grayscale image. Then blurring the (ROI) by Gaussian blur to reduce the objects that have high frequency but not the target. Notice that in this phase the algorithm will fail if there is any vibration for the camera.

4.3. Hand region segmentation
This phase is important in any system to hand gesture recognition and help in enhancing the performance of the system by removing unwanted data in the video stream. In general, there are two methods to detect the hand in the image. The first method depends on Skin-Color, it is simple but effected by the light conditions in the environment and the nature of background, also suffer from clutter due to existing objects such as the face or arm having the same color of hand. This method can be used as a threshold technique that exploits the color distribution map information in a suitable space of color. The color of the skin varies significantly among people, especially among people of different
races in addition to the impact of lighting. To solve this problem, the researchers suggested relying on the chromaticity of the skin because it is the same for all races and contains important information, in contrast to the luminance, which is heavily influenced by lighting [18]. Thus, it is possible to say that the best space of color to detect the color of the skin is what separates the luminance from the chromaticity.

The second method does not depend on Skin-Color, but on the shape of hand and benefits from the principle of convexity in the detection of the hand. In this paper, using contour analysis that depends on the shape and to solve the problems of Skin-Color detection.

There are several methods or techniques used to extract the hand region from the image can be summarized as:

1. Edge-Detection.
2. RGB values because of the values of RGB for the hand different from the background of the image.
3. Subtraction of background.

This work used the background subtraction technique to eliminate all objects that be static and considered it as a background and then to separate the hand from the background. This technique needs to determine the background that can be obtained by benefit from a running average principal. The initial background is computing from making the system focus on a certain scene for at least 30 frames to obtain the average as in the equation (1). After determining the initial background, we put the hand front the camera, then computes the absolute difference between the initial background and the current frame that contains the hand as a foreground object as given in the equation (2). Finally, calculating the running average to update the background using equation (3).

\[
Average_{(x,y)} = \frac{1}{N} \sum_{i=1}^{N} Frame(i)_{(x,y)}
\]  

(1)

\[
Foreground-object = \left| Frame_{(x,y)} - Average_{(x,y)} \right|
\]  

(2)

\[
Average_{(x,y)} = (1 - \alpha) Average_{(x,y)} + \alpha Frame_{(x,y)}
\]  

(3)

where \( Frame_{(x,y)} \) is a source image (the current frame) may be one or three channels and 8-bits or 32-bits floating-point, \( Average_{(x,y)} \) is the destination image (the average background) contains the same channels in the source image and 32-bits or 64-bits floating point. Alpha is a weight of the source image and can be considered as a threshold to determine the time for computing the running average over frames. Thus find the background and then compute the difference all that called the subtraction background.

In general, the background subtraction technique faces many challenges such as interference between objects, noise from the motion camera, shadow, and change of illumination, all these challenges are taken into account in this paper. The next step is thresholds the image that output from background subtraction and the result will be the only hand with white color and the rest image with black color. The threshold process is important and must be done before finding the contours to achieve high accuracy. Mathematically can represent the threshold principle as follows:

\[
f(x) = \begin{cases} 
1, & x \geq \text{threshold} \\
0, & x < \text{threshold}
\end{cases}
\]  

(4)

where \( f(x) \) is the intensity of the pixel.

All of the above processes are called Motion-Detection. Figure (4) is shown the output of the hand region segmentation process. Finally, perform a chain of morphological processes such as Erosion and Dilation to remove any small regions of noise.
Figure 4. The Output of Hand region segmentation process.

4.4. Contour-Extraction
The contour can be defined as the object’s boundary or outline (hand in our case) that be located in the image. In other words, the contour is a curve connecting points that have the like color value and is a very important feature in shape analysis, objects detection and recognition process.

4.5. Features extraction and recognition
After extracted a contour of hand as a feature, now we turn to the second part of the research, which is how we determine the number of fingers. From the number of fingers can recognize hand gestures, and for performing this task merged two methods, one used Convex Hull to locate the extreme points (top, bottom, left and right) and the other depends on Convexity Defects.

4.5.1. Convex hull method. The convex hull is the group of points that surrounding the region of hand, as shown in Figure (5).

Figure 5. Convex Hull.

Here, we must clarify the principle of the Convex Set, which means all lines between any two points within the hull are entirely within it. From extreme points can compute the center of palm, see Figure (6).
The next step is to draw a circle about the fingers, it’s the center point is the center of the palm and seventy percent from the length of maximum Euclidean distance between the palm’s center and extreme points represent the radius. After that using the bitwise AND operation between the circle from the previous step and the threshold image, slices of fingers that result from this operation can use to compute the number of hand fingers. After determined the gesture based on the number of hand’s fingers, the corresponding operation is performed. In fact, the process of distinguishing the hand gesture is a dynamic process. After performing the required instruction from the gesture, return to the first step to take another image to be processed and so on.

4.5.2. Convexity defects method. Figure (5) explains the Convex Hull and for preserving the feature of convexity formation the defects as shown in Figure (7). There is a defect when the object's contour is away from the Convex Hull to the object itself. Convexity defect is a vector contains three points (start, end, farthest) and the approximate distance between the farthest point and convex hull as shown in Figure (8). After found the defects must obtain the angle between the two fingers to determine the finger is held up or not. From the triangle formed by points (start, end, and farthest) can compute the angle. Then used the Euclidean Distance equation to find the length of the lines of the triangle as:

\[ A = \sqrt{(start[0] - end[0])^2 + (start[1] - end[1])^2} \] (5)

\[ B = \sqrt{(start[0] - farthest[0])^2 + (start[1] - farthest[1])^2} \] (6)

\[ C = \sqrt{(end[0] - farthest[0])^2 + (end[1] - farthest[1])^2} \] (7)
After that, and by cosine base can find the angle farthest as follow:

\[ A^2 = B^2 + C^2 - 2BC \cos \text{Farthest} \]  

(8)

Then:

\[ \text{Farthest} = \cos^{-1}(B^2 + C^2 - A^2/2B*C) \]  

(9)

If the farthest angle is less than or equal to 85° then the two fingers are considered held up. This can use to count the fingers from two to five by:

Number of fingers = number of defects + 1

(10)

5. Results and analysis

This paper recognized sixteen gestures as shown in Figure (9) and Figure (10). The first method formed six gestures, but the recognition of the Five gesture is not perfect. After the combined Convex Hull method with the Convexity Defects method, the recognition of the Five gesture became perfect and added another ten gestures as shown in Figure (9). The parameters that used to recognize last ten gestures are the number of defects, the number of fingers, the distance between the start point and the endpoint, the distance between the endpoint and the farthest point, the distance between the start point and the farthest point, coordinates of extreme points, the farthest angle and the distance between extreme points and coordinates of the center point.

**Figure 8.** Components of Convexity Defects.

**Figure 9.** The shape of six gestures.
The sixteen gestures recognized in this paper are used to generate the actions of the mouse and used to open various applications such as Microsoft Office (Word, Excel, PowerPoint), VLC Media Player, Google Chrome, Internet Explorer, etc. The system applied in different conditions, it has given the best results when the background is clear and the light is medium, the rate of recognition reached 100%. Table 1 shows the experimental results of the proposed system with threshold values (50 when background brighter than skin color, 90 when skin color brighter than the background) and the alpha value of 0.5. The performance of the system is given in Chart 1. The results of the system are shown that the rate of recognition is 97.5% and this result is considered very good compared with other research papers as explained in Table 2.
Table 1. Hand Gesture Recognition Rate.

| Gesture name       | No. of input samples | No. of recognized samples | Recognition rate (%) | Total recognition rate(%) |
|--------------------|----------------------|---------------------------|-----------------------|---------------------------|
| Zero               | 10                   | 10                        | 100                   |                           |
| One                | 10                   | 10                        | 100                   |                           |
| Two                | 10                   | 10                        | 100                   |                           |
| Three              | 10                   | 10                        | 100                   |                           |
| Four               | 10                   | 10                        | 100                   |                           |
| Five               | 10                   | 10                        | 100                   |                           |
| Rock and roll      | 10                   | 10                        | 100                   |                           |
| Engy               | 10                   | 10                        | 100                   |                           |
| Love               | 10                   | 10                        | 100                   |                           |
| Move left          | 10                   | 10                        | 100                   |                           |
| Move right         | 10                   | 10                        | 100                   |                           |
| Rock               | 10                   | 10                        | 100                   |                           |
| Ok                 | 10                   | 9                         | 90                    |                           |
| Pinky              | 10                   | 9                         | 90                    |                           |
| Shocker            | 10                   | 10                        | 100                   |                           |
| Palm               | 10                   | 8                         | 80                    |                           |

Figure 11. Performance Results

Table 2. The recognition rate of the proposed method versus some previous methods.

| Method            | No. of gestures | Recognition rate (%) |
|-------------------|-----------------|----------------------|
| Proposed method   | 16              | 97.5                 |
| [15]              | 4               | 88.7                 |
| [20]              | 6               | 71.30                |
| [20]              | 6               | 86.90                |
| [21]              | 10              | 94.86                |
| [21]              | 10              | 92.07                |
| [19]              | 8               | 96.25                |
| [9]               | 10              | 94.3                 |
6. Conclusion
For a long time, the problem of distinguishing gestures was important in computer vision because of the challenge of extraction of the target object such as hand from a background that has been cluttered and all that in real-time. The human when looking to a certain image can easily detecting what is inside it but that very difficult for the computer if it looks to the same image because it deals with the image as a Matrix with three dimensions.

In this paper, we got the same results when used right or left hand. The system provided used only bare hand and webcam of Laptop so it is very flexible for the user. The system does not need a database, but it directly distinguishes the gesture and this achieves the speed of the system. The contribution in this paper is combined two methods, Convex Hull and Convexity Defects to recognize sixteen hand gestures. In the future and to enhance the system can use both hands instead of using only the right hand, and that will increase the number of gestures. The experimental results showed that the best rate of recognition is when the background is clear and the light is medium, so these limits must be addressed in the future in order to increase the accuracy of the system.

7. References
[1] Ma, B., Xu, W., & Wang, S. 2013 A robot control system based on gesture recognition using Kinect. TELKOMNIKA Indonesian Journal of Electrical Engineering, 11(5), pp. 2605-2611.
[2] Lamar, M. V., Bhuiyan, M. S., & Iwata, A. 2000, May T-CombNET-A Neural Network Dedicated to Hand Gesture Recognition. In International Workshop on Biologically Motivated Computer Vision (Vol.1811, pp. 613-622). Springer, Berlin, Heidelberg
[3] Patel, N., & Dr. He, S. 2018, June A Survey on Hand Gesture Recognition Techniques, Methods and Tools. International Journal of Research in Advent Technology (IJRAT), 6(6), pp.939-957
[4] Bilal, S., Akmelawiati, R., El Salami, M. J., & Shafie, A. A. 2011, May Vision-based hand posture detection and recognition for Sign Language—A study. In Mechatronics (ICOM), 2011 4th International Conference On (pp. 1-6). IEEE
[5] Rautaray, S. S., & Agrawal, A. 2015 Vision-based hand gesture recognition for human-computer interaction: a survey. Artificial intelligence review, 43(1), pp.1-54
[6] Garg, P., Aggarwal, N., & Sofat, S. 2009 Vision-based hand gesture recognition. World Academy of Science, Engineering and Technology, 49(1), pp. 972-977
[7] Dipietro, L., Sabatini, A. M., & Dario, P. 2008 A survey of glove-based systems and their applications. IEEE Trans. Systems, Man, and Cybernetics, Part C, Vol.38 (4), pp.461-482
[8] LaViola, J. 1999 A survey of hand posture and gesture recognition techniques and technology, Master Thesis, Brown University, Providence, RI, 29
[9] Lai, H. Y., Ke, H. Y., & Hsu, Y. C. 2018 Real-time Hand Gesture Recognition System and Application. Sensors and Materials, 30(4), pp.869-884
[10] Hasan, M. M., & Mishra, P. K. 2012 Hand gesture modeling and recognition using geometric features: a review. Canadian Journal on Image Processing and Computer Vision, 3(1), pp. 12-26
[11] Lamberti, L., & Camasta, F. 2011, September Real-time hand gesture recognition using a color glove. In International Conference on Image Analysis and Processing (pp. 365-373). Springer, Berlin, Heidelberg
[12] Dhawan, A., & Honrao, V. 2013, June Implementation of hand detection based techniques for human-computer interaction. International Journal of Computer Applications, 72(17), pp. 0975–8887
[13] Yewale, S. K., & Bharne, P. K. 2011, April Hand gesture recognition using different algorithms based on artificial neural network. In 2011 International conference on emerging trends in networks and computer communications (ETNCC) (pp. 287-292). IEEE
[14] Nancy, Sekhon, G.S., Kashi, G. (2012, June). An analysis of hand gesture recognition technique using finger movement detection based on color marker. International Journal of Computer Science & Communication, 3(1), pp. 129-133
[15] Ahmed, T. 2012 A neural network based real-time hand gesture recognition system. *International journal of computer applications* (0975 – 8887), 59(4).

[16] Badgujar, S. D., Talukdar, G., Gondhalekar, Kulkarni, O. S. Y. (2014, February). Hand Gesture Recognition System. *International Journal of Scientific and Research Publications*. 4(2), pp. 2250-3153.

[17] Nagarajan, S., Subashini, T. S., & Ramalingam, V. 2013 Vision-Based Real-Time Finger Counter for Hand Gesture Recognition. CPMR-IJT: *International Journal of Technology*. 2(2).

[18] Singha, J., & Das, K. 2013 Hand gesture recognition based on Karhunen-Loeve Transform. Mobile and embedded technology, International Conference.

[19] Chaikhumpha, T., & Chomphuwiset, P. 2018, January Real-time two hand gesture recognition with condensation and hidden Markov models. In *2018 International Workshop on Advanced Image Technology (IWAIT)* (pp. 1-4). IEEE

[20] Badi, H. (2016). Recent methods in vision-based hand gesture recognition. *International Journal of Data Science and Analytics*, 1(2), pp.77-87

[21] Nasr-Esfahani, E., Karimi, N., Soroushmehr, S. M., Jafari, M. H., Khorsandi, M. A., Samavi, S., & Najarian, K. (2016). Hand Gesture Recognition for Contactless Device Control in Operating Rooms. *arXiv preprint arXiv:1611.04138*. 