Vision-Based Real-Time Human–Computer Interaction on Hand Gesture Recognition

Poorvika Singh Negi, Riya Pawar and Roshan Lal

Abstract Gesture recognition technology has evolved greatly over the years. It is a field in language innovation and development with the aim of deciphering human motions by the use of certain algorithms. In this age of technology, we have fabricated boundless techniques and subsequently seen their drawbacks making us aware of our own limits in terms of speed and naturalness of the human body. The intellect and inventiveness of human beings have led to the development of many tools, gesture recognition technology being one of them. They help us extend the capabilities of our senses by combining natural gestures to operate technology, thereby reducing human efforts and going beyond human abilities. Gestures can be viewed as a route for PCs to start to comprehend human non-verbal communication, in this manner fabricating a more extravagant extension among machines and people than crude content UIs or GUIs, which still breaking point most of contribution to console and mouse and connect normally with no mechanical gadgets. Utilizing the idea of motion acknowledgment, it is conceivable to point a finger now will move as needs be. With the expeditious inventions of three-dimensional applications and the upcoming hype of virtual environments in systems, there is a need of new devices that can sustain these interactions. The advancement of user interface molds and develops the human-computer interaction (HCI). The paper aims to present a review of vision-based hand gesture recognition techniques for HCI and using it to tally finger count using the same.

Keywords Gesture · Computer vision · Real-time video · Human–computer interaction (HCI) · Contour extraction · Convex hull · Euclidean distance · Region of interest (ROI)

P. S. Negi (✉) · R. Pawar · R. Lal
Amity University, Noida, UP, India
e-mail: negikpoorvika@gmail.com

R. Pawar
e-mail: riyapanwar1999@gmail.com

R. Lal
e-mail: rchhokar@amity.edu

1 Introduction

Gestures are an important and striking means of communication. A natural movement caused by a body or a part of the body does not have a significant meaning and therefore differentiates gestures as the later transmits significant meaning to the observer that is receiving them [1]. Human and computing device interaction, which is inspired by the natural human-to-human interactions, can be achieved by hand gesture technology [2].

Wrapping this technology with computer vision makes it more dynamic and gives a detailed interpreted approach to gesture recognition technology [3, 4]. It is an interdisciplinary science, which is making its way in innumerable sectors, from retail and security to automotive, health, agriculture, and banking industry. From a technical point of view, we try to automate the tasks that the human vision system can do [5]. In simple words, it can be explained as when a computer and/or machine is given the capability of sight.

Picking up the accelerated speed that technology is attainted today, when vision is hooked with certain algorithms, it transforms and takes a new step into another highly advanced field like machine and deep learning [6]. It grants the machine power of recognition of images, interpretation of images and solutions, and even learns in some cases [7].

This project gives a detailed analysis and review of gesture recognition technology when approached by computer vision, focusing on creating hand gesture recognition and concepts like background subtraction, motion detection, thresholding, and contours [8]. These concepts are implemented using OpenCV and python and show how to segment hand region from a real-time video sequence. Furthermore, the program recognizes the number of fingers shown in the real-time video sequence and shows the output on the screen.

2 Literature Review

The ever-changing nature of data requires us to have different kinds of methodology for recognizing and deciphering a signal in numerous ways [9]. However, many methods depend on key pointers spoke to the three-dimensionally arranged framework [10]. A signal can be identified with precision in view of the overall movement of the gesture.

The first step to translate development of the body will be to order them according to their basic properties and the message that the development may express [11]. Considering the conversations that take place with the help of gestures, every word or expression is communicated via them. Quek proposed in "Towards a Vision Based Hand Gesture Interface" a scientific classification that goes hand in hand with human–computer interaction (HCI) [12]. He divides them into the following so as to broadly categorize all kinds of motions:

2.1 Manipulative Gesture System

These kinds of system follow the traditional approach that was given by Richard Bolt, which is commonly known as "Put-That-There Approach." Here, the system permits direct communication and allows the user to interact with big display objects that are in motion around the display screen [12]. The major characteristic of this system is that the gesture and the entity being controlled are coupled together and have a tight response in between each other. It is very similar to the direct interfaces that have been manipulated. However, the only difference in both is that there is a "device" present in this gesture system [3].

2.2 Semaphoric Gesture System

This gesture system approach may be called "communicative" as the gestures here can be said to be equivalent to numerous gestures that are used to communicate with a machine. Each gesture/movement/pose may have a particular and designated significance [12]. However, it is important to note that unlike sign language, which has their own syntax and dynamics, the semaphoric gesture system approach only comprises of isolated symbols [13].

Semaphores amount to a very small portion in the amount of gestures that are used in our day-to-day lives. But it is to be noted that they are frequently used in literature as they can most easily be achieved [14].

2.3 Conversational Gesture System

The gestures performed in due course of time by humans in their day-to-day life are known as conversational gestures. They come to us naturally, and the person making these gestures gives not a lot of thought [2, 12]. They do not have a particular or specifically fixed meaning and can be used in one more context unlike the semaphoric gesture system where all the signs and gestures have fixed syntax, grammar, and meanings. They are isolated symbols and are generally accompanied by language or speech [1]. Even though they are not consciously constructed by the human mind, they can still be determined by disclosure text, personal style, culture, social presence, etc.

3   Methodology

In this paper, a hand signal is perceived from a real-time video succession. To perceive the motions from a live arrangement, the first step would be to take out the hand area distinctly, evacuating the undesirable bits in the video succession [8]. In the wake of portioning the hand district, we at that point include the fingers appeared in the video succession to teach the robot dependent about the finger tally.

3.1   Hand Segmentation

The initial phase close by motion acknowledgment is clearly to discover the hand locale by taking out the various undesirable segments in the video arrangement [15]. After an amount of processing has already taken place, it is determined what areas of the image or what particular pixels are relevant for the process to move further.

3.2   Background Subtraction

In the first place, we need an effective technique to separate closer view from foundation. To do this, we utilize the idea of running midpoints [14]. We make our framework to investigate a specific scene for 30 outlines.

During this period, we register the running normal over the present casing and the past edges. By doing this, we basically make the framework aware of the foundation [16] (Fig. 1).

3.3   Threshold

To distinguish the hand district from this distinction picture, we have to limit the distinction picture, with the goal that lone our hand area winds up noticeable and the various undesirable areas are painted as dark [16]. This is called motion detection.

Threshold is the assignment of pixel powers to 0s and 1s based on a specific edge level with the goal that our object of intrigue alone is caught from a picture [6, 15].

3.4   Contour Extraction

The next phase is the contour extraction. In simple words, it is basically a technique with the help of which we extract the boundary of the required part in the digital
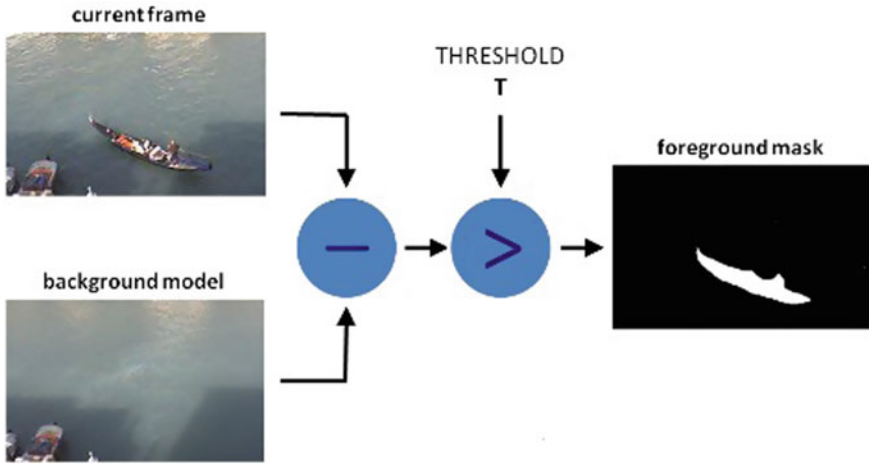
Fig. 1 Procedure to extract foreground mask. Sourced from https://gogul.dev/software/hand-gesture-recognition-p1

image in which we are applying it [9]. It gives us information about the shape of the detected part [13]. Characteristics are examined after the extraction and then further used as features in pattern classification (Fig. 2).

Fig. 2 Picture acquired after segmentation, subtraction, threshold, and contour extraction
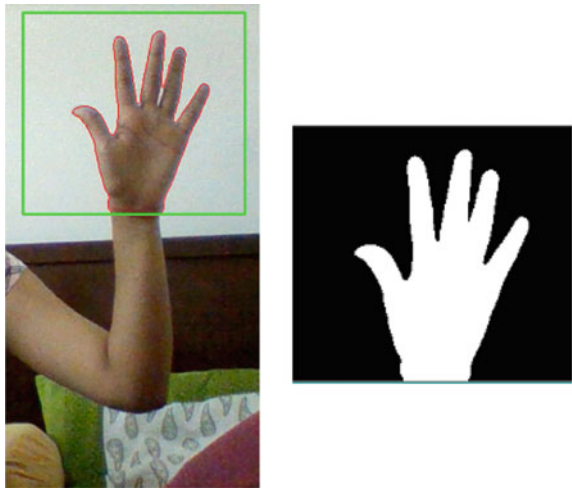
3.5 Finger Tally

The hand region has been obtained from the real-time video sequence (by the steps written above). To tally the fingers, the foremost requirement is the accessibility of a webcam or a camera attached to the system [2]. We have obtained the segmented hand region by assuming it as the most important contour (i.e., contour with the maximum location) inside the frame, and therefore, it is pertinent to that the hand occupies majority of the area inside the frame [17].

The next step is to construct an intermediary circle around the palm [13]. This is done by first detecting and computing the extreme points of the convex hull of the obtained hand region, taking them as the points joining to be the perimeter of the circle. Taking radius as the maximum Euclidean distance, the circle is constructed and henceforth we apply bitwise AND on the region of interest (ROI) and the frame [4] (Table 1).

| Table 1 Algorithm to obtain finger tally count		
		Step 1: obtain hand region
		Step 2: detect convex hulls and obtain extreme points
		Step 3: find center of palm
		Step 4: construct circle
		Step 5: perform bitwise AND between ROI and frame
		Finger tally obtained

4 Result

When the hand is brought within the bounding box such as it is occupying the majority region, the hand is detected and a red dotted trace outlining the hand appears on the real-time video. As the fingers are gestured in the green bounding box, finger count is displayed on the top-left corner of the window in red color (Figs. 3, 4, and 5).
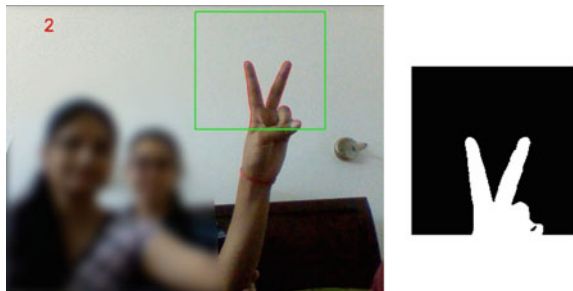

Fig. 3 Video feed and threshold for finger tally with output 0
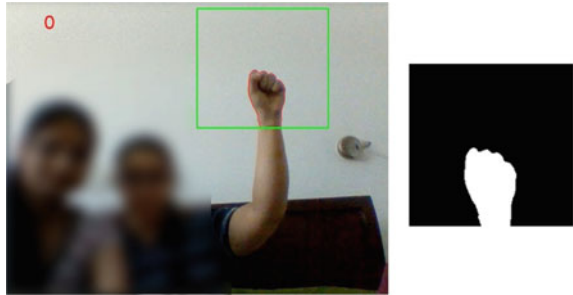
Fig. 4 Video feed and threshold for finger tally with output 1
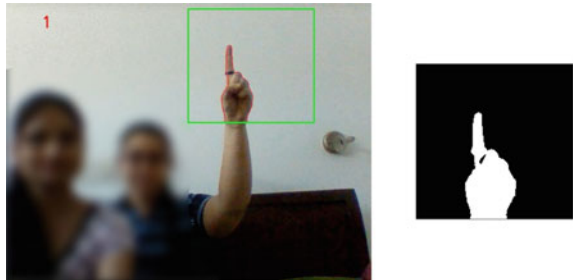
Fig. 5 Video feed and threshold for finger tally with output 2

5 Conclusion

In this paper, we presented a review of vision-based hand gesture recognition techniques for HCI. With the help of different gesture systems, we classified the basic kinds of gestures. After reviewing the basic comebacks and difficulties that are faced in gesture recognition technology, we approached the same using computer vision. With the help of segmentation, background subtraction, motion detecting, thresholding, and contour extraction, we were successfully able to detect and note down the characteristics of a hand in the real-time video.

Using the algorithm that was discussed in the paper, we were successfully able to detect the number of fingers or finger count and it was displayed on the screen.

References

1. Murao K, Terada T, Yano A, Matsukura R (2011) Evaluating gesture recognition by multiple-sensor-containing mobile devices. In: 15th Annual International Symposium on Wearable Computers, San Francisco, USA
2. Sarkar AR, Sanyal G, Majumder S (2013) Int J Comput Appl 71(15):0975–8887
3. Saeed A, Bhatti MS, Ajmal M, Waseem A, Akbar A, Mahmood A (2013) Android, GIS and web base project, emergency management system (EMS) which overcomes quick emergency response challenges. In: Advances in intelligent systems and computing, vol 206. Springer, Berlin
4. Nanaware T, Sahasrabudhe S, Ayer N, Christo R (2018) Finger spelling-Indian sign language. In: IEEE 18th international conference on advance learning technologies
5. Chiang T, Fan C-P (2018) 3D depth information based 2D low-complexity hand posture and gesture recognition design for human computer interactions. In: 3rd international conference on computer and communication systems
6. Lin M, Mo G (2011) Eye gestures recognition technology in human-computer interaction. In: 4th international conference on biomedical engineering and informatics (BMEI), Shanghai, China
7. Geer D (2004) Will gesture recognition technology point the way. Computer 37(10):20–23
8. Phade GM, Uddharwar PD, Dhulekar PA, Gandhe ST (2014) Motion estimation for human-machine interaction. In: International symposium on signal processing and information technology (ISSPIT)
9. Yu C, Wang X, Huang H, Shen J, Wu K (2010) Vision-based hand gesture recognition using combinational features. In: Sixth international conference on intelligent information hiding and multimedia signal processing
10. Konwar P, Bordoloi H (2015) An EOG signal based framework to control a wheel chair. IGI Global, USA
11. Murao K, Yano A, Terada T, Matsukura R (2012) Evaluation study on sensor placement and gesture selection for mobile devices. In: Proceedings of the 11th international conference on mobile and ubiquitous multimedia. New York, USA
12. Quek FKH (1994) Toward a vision-based hand gesture interface. Virtual Reality Software and Technology, pp 17–31
13. Chayapathy V, Anitha GS, Sharath B (2017) IOT based home automation by using personal assistant. In: International conference on smart technology for smart nation
14. Shukla J, Dwivedi A (2014) A method for hand gesture recognition. In: Fourth international conference on communication systems and network technologies

15. Yilmaz A, Javed O, Shah M (2006) Object tracking: a survey. In: ACM computing surveys, Dec 2006
16. Bao PT, Binh NT (2009) A new approach to hand tracking and gesture recognition by a new feature type and HMM. In: Sixth international conference on fuzzy systems and knowledge discovery
17. Chaikhumpha T, Chomphuwiset P (2018) Real-time two hand gesture recognition with condensation and hidden Markov models. In: International workshop on advanced image technology (IWAIT)