

Research Article

Using Depth Cameras for Recognition and Segmentation of Hand Gestures

Khalid Twarish Alhamazani ¹, **Jalawi Alshudukhi** ¹, **Talal Saad Alharbi** ¹,
Saud Aljaloud ¹ and **Zelalem Meraf** ²

¹University of Ha'il College of Computer Science and Engineering, Department of Computer Science, Hail, Saudi Arabia

²Department of Statistics, Injibara University, Injibara, Ethiopia

Correspondence should be addressed to Zelalem Meraf; zelalemmeraf@inu.edu.et

Received 21 November 2021; Accepted 8 December 2021; Published 18 December 2021

Academic Editor: Palanivel Velmurugan

Copyright © 2021 Khalid Twarish Alhamazani et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, in combination with technological advances, new paradigms of interaction with the user have emerged. This has motivated the industry to create increasingly powerful and accessible natural user interface devices. In particular, depth cameras have achieved high levels of user adoption. These devices include the Microsoft Kinect, the Intel RealSense, and the Leap Motion Controller. This type of device facilitates the acquisition of data in human activity recognition. Hand gestures can be static or dynamic, depending on whether they present movement in the image sequences. Hand gesture recognition enables human-computer interaction (HCI) system developers to create more immersive, natural, and intuitive experiences and interactions. However, this task is not easy. That is why, in the academy, this problem has been addressed using machine learning techniques. The experiments carried out have shown very encouraging results indicating that the choice of this type of architecture allows obtaining an excellent efficiency of parameters and prediction times. It should be noted that the tests are carried out on a set of relevant data from the area. Based on this, the performance of this proposal is analysed about different scenarios such as lighting variation or camera movement, different types of gestures, and sensitivity or bias by people, among others. In this article, we will look at how infrared camera images can be used to segment, classify, and recognise one-handed gestures in a variety of lighting conditions. A standard webcam was modified, and an infrared filter was added to the lens to create the infrared camera. The scene was illuminated by additional infrared LED structures, allowing it to be used in various lighting conditions.

1. Introduction

The effectiveness of the human-computer interaction is directly influenced by the way the interface is designed. With the rise of personal computers in the 1980s, HCI arose. Computers were no longer being created just for professionals, and HCI's purpose was to make all computer interactions simple and efficient for a wide range of users with varying skill levels.

Human-computer interaction (HCI for its acronym in English) is a prevalent concept among large companies. A more intuitive relationship between man and computer is sought to apply it in different technological devices. It is considering end-user satisfaction by reducing their effort to perform tasks. Projects have been worked on where communication is carried

out through body gestures; Microsoft is one of the companies that made this possible with one of its projects called Kinect; another case is that of Nintendo with the Wii console. That is why, contributing to the improvement of HCI and the new trends of creating interaction with body gestures, the search for interaction with the user is proposed where it is required to identify a hand using infrared light as a critical tool [1].

Infrared light provides information that cannot be obtained from visible light. All bodies tend to emit infrared radiation, which depends directly on the temperature in which the body is located, and this is one of the advantages that are taken advantage of within the developments of computer vision since the image capture provides clear images, with really striking lighting that allows identifying bodies in the scene. An infrared camera is a device that, from the middle infrared emissions of

the electromagnetic spectrum of the detected bodies, forms luminous images visible to the human eye. The risk of common eye illnesses has been linked to high ambient temperatures. This type of infrared shot is associated with night shots or the possibility of seeing in very dark situations [2].

1.1. Review of Literature. Hameed et al. [3] proposed developing a gesture recognition system to communicate through a more natural human-computer interaction. The main objective is to create a robust and efficient segmentation algorithm based on colour spaces and morphological processing necessary for skin colour detection, image background removal, and variable lighting conditions. In work, the OpenCV library was used for monitoring and a perimeter travel algorithm to detect the hand contour.

According to [4], detection of gestures is one of the important parts of HCI, which is why he proposes a work in which the detection of gestures is done through pre-processing of images to reduce noise in addition to using support vector machines (SVM) for the detection and extraction of the region where the hands are located using prominent characteristics with an appearance-based approach.

This segmentation technique is based on combining two unsupervised clustering approaches such as K -means and expectation maximization. The experimental results showed that this technique succeeds in correctly segmenting the hand from the body, the face, the arm of a person, and other elements found in the image, which was taken with variable lighting conditions in real-time.

Ito et al. [5] developed a method of obtaining hand characteristics from image sequences, specifically when a person performs the signs of the Japanese sign language to recognise the words of a said language in a complex background. For recognition, the hidden Markov model (HMM) technique is used using six features of the face and hands. The results presented indicate that the objective of monitoring the face and hands is met.

On the other hand, in [6], a method of predicting the poses of an articulated hand is presented in real-time with a depth camera to carry out the interaction in an environment of mixed reality and to study the effects of models of real and virtual articulated hands in a simulator. Random decision forests were used to carry out the recognition, which proved to have better results in real-time applications, avoiding the typical errors in the use of these technologies and showing low consumption of computational resources and high precision.

In this work, an infrared camera is built from small modifications at the hardware level of a conventional webcam, achieving a clean image, and through the segmentation of the scene, the hand is identified, finally leading to a gesture recognition process. The contribution of this work can be determined as obtaining a device that allows images to be captured with the characteristics mentioned earlier and with which the possibility of being used in projects that require capturing scenes. It should be noted that the tests are carried out on a set of relevant data from the area. Based on this, the

performance of this proposal is analyzed about different scenarios such as lighting variation or camera movement, different types of gestures, sensitivity, or bias by people, among others.

2. Methodology

For the development of this work, it was decided to adopt the improved cascade methodology used in software development [6] because this methodology establishes the stages of the software life cycle so that the beginning of each phase must wait. However, the end of the previous one allows the improvement of prior stages, if necessary. In Figure 1, the scheme of the methodology that includes three steps is shown.

The first stage, named “hardware adaptation,” refers to all the activities to obtain the camera with the necessary modifications and the best alternative for capturing the desired image to continue with the process. In the second stage, “image processing and feature extraction,” essential aspects are defined for the features to be compared and the structure of the library to be built. The third stage, “gesture recognition and implementation,” refers to how the training is achieved to detect gestures and implement the library in an application.

2.1. Hardware Adaptation. This section presents the hardware structure and its modifications to obtain an infrared camera with the necessary characteristics for the project. Figure 2 shows a diagram of the camera and the essential additions to convert it into an infrared camera.

First, the web camera (Figure 2(a)) model COM-105 with a 480-pixel CMOS sensor is modified so that the ICR filter is eliminated following the procedure explained in (WANG 2013). Subsequently, to make adequate lighting of the scene that would be evaluated during the tests, building an array of LEDs (Figure 2(c)) is necessary.

In a second phase, this structure is built using an arrangement of 10 LEDs model TN130BF, whose greater illumination amplitude angle allows the objects to be segmented more completely, unlike other diode models.

Finally, some materials that allow filtering the infrared light are used to determine the elements present in the image (Figures 2(a) and 2(b)). The materials used were selected considering the wavelength captured by each of the materials, carrying out experimental tests to determine the one that offered the appropriate capture of the hand and the gestures performed. At the end of these tests, the negative of a roll photographic film was selected as the material (Figure 3), allowing observing elements with a wavelength between 800 nm and 1000 nm.

2.2. Image Processing and Feature Extraction. The term “digital image processing” refers to the use of a computer to process digital images. We can also call it the application of computer algorithms in order to obtain a better image or extract useful information. After the camera had been modified, it had to be checked to see whether the acquired pictures had enough information to recognise one-handed

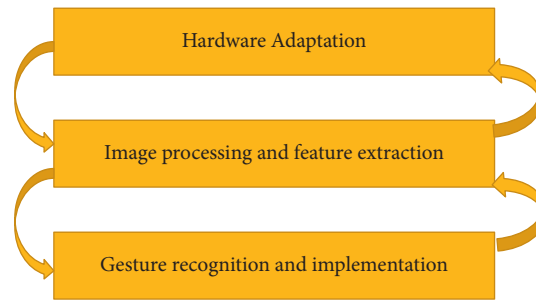


FIGURE 1: Scheme of the proposed methodology.

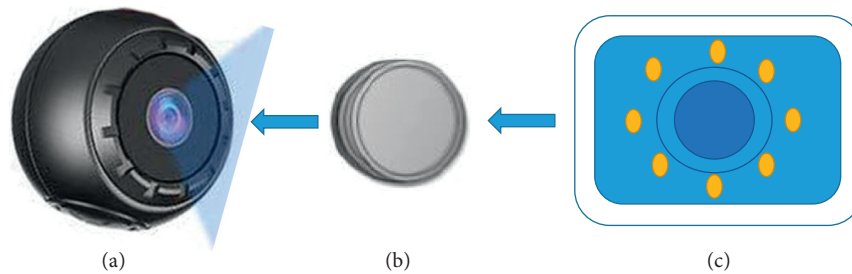


FIGURE 2: Chamber schematic and additions.

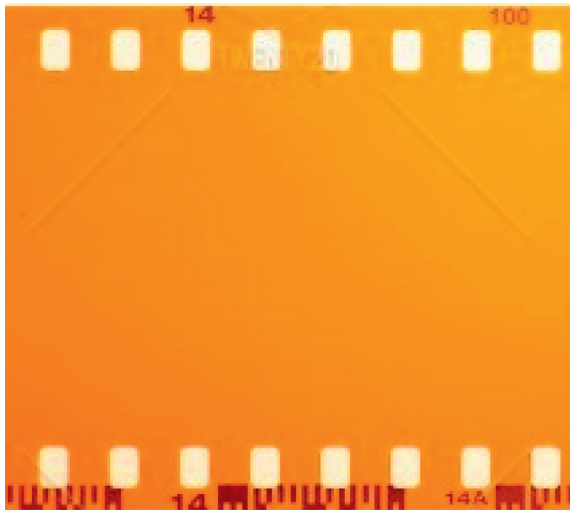


FIGURE 3: Photographic film negative strip.

motions. A library of functions based on the C++ language and the OpenCV library [7] were created to carry out the verification.

The first step to carry out is the hand segmentation in the foreground from the rest or background of the captured image. The way selected to perform this segmentation is based on threshold segmentation where, based on values established by a person, a colour image was obtained by the camera constructed through a mask that verifies pixel by pixel, if the numerical value of each one of these pixels is in the range that defines a hand. Otherwise, it is determined that that pixel does not belong to the hand. Figure 4 shows the process for establishing the threshold for a colour channel found in the images.

Because one of the characteristics of the images captured by an infrared camera is that the objects reflect this type of light, the use of this channel was analysed for a program to transform the information in it into another image at a scale of greys to support better segmentation of the hand in the image. This transformation offered an improvement in segmentation, which is why it was selected to continue using the grayscale. In image five, you can see the image resulting from this process.

Since the hand has been segmented, the defining characteristics must be extracted. These characteristics will be obtained from the analysis of the contour of the hand and the location of the centroid of the hand.

The hand contour is obtained as a sequence of points using the algorithm developed by Satoshi in [7] and is implemented in the OpenCV library. Once received, the convex envelope (larger polygon that surrounds a figure in the image) and its defects (deep points between the convex envelopes) must be found [8–14]. This allows determining the space occupied by the hand and locating the fingertips, as shown in Figure 5. In the figure, you can see the beginning and end of a defect that corresponds to the union of the space between the fingers of the hand and their tip.

After the location of the largest polygon, which corresponds to the hand with the fingers included, the centroid of the hand is determined using the detected polygon (Figure 6).

With the above, it is determined that the characteristics to be used to identify the gestures would be defined by a defect and the centroid, as defined in Table 1.

With the above, it is established that 20 characteristics are used to represent an extended hand to which the width and height of the hand are added to make a total of 22 elements to analyse (Figure 7).



FIGURE 4: Assignment of the threshold for the channel with the colour red and segmented with that information.



FIGURE 5: Threshold assignment for grayscale and segmented with such information.

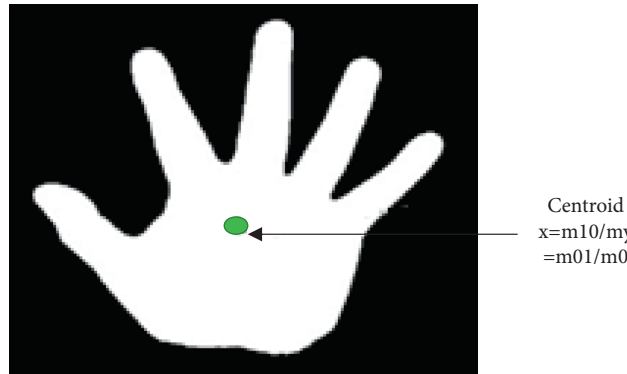


FIGURE 6: Location and description of the convex envelope and its defects.

TABLE 1: Description of extracted characteristics.

Feature in the picture	Feature extracted from the image
Default	Distance from start to finish Starting distance to the deepest point Distance from the end to the deepest point
Centroid	Distance from start to the centroid Distance from end to the centroid

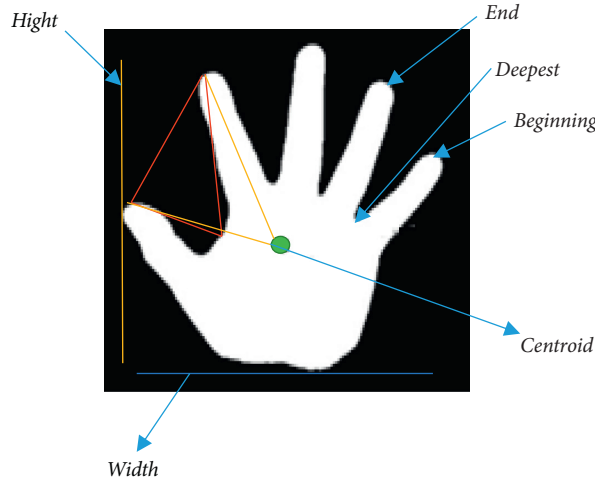


FIGURE 7: Description of extracted features.

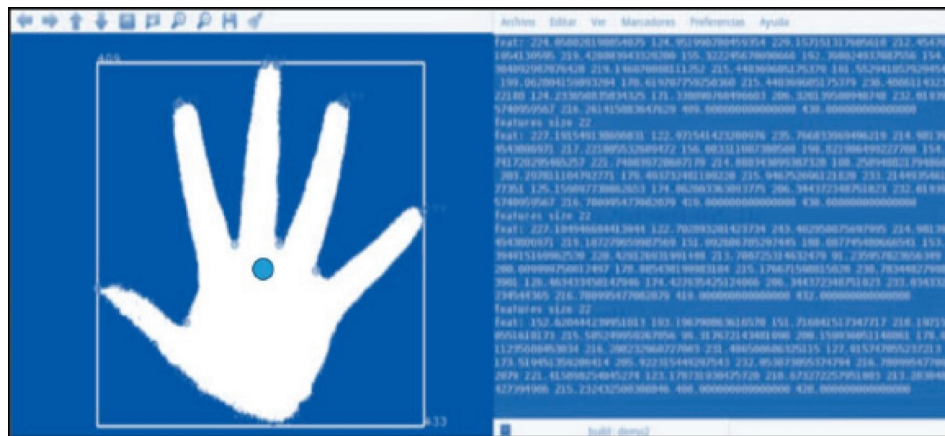


FIGURE 8: Hand identification with five sides extended and features removed.

3. Results

It can be concluded that both the built camera and the implemented library can support the detection and recognition of hand gestures. This was achieved using the technique of vector support machines (SVM) [15], which learns to classify data into two different classes. An implementation of SVM is included in the OpenCV library with which it was possible to use it.

To perform the training of the machines, the extraction of characteristics of various shots taken with the developed library was executed, and each gesture shot was labelled as five fingers and four fingers. Examples of the parts are presented in Figure 8.

For each gesture of extended fingers, 55 samples were used, with which the machines were trained to later verify with images captured directly from the camera turned on. Each gesture is checked separately with another motion where the fingers are not separated, as shown in Figure 8.

The results obtained allowed us to observe that the development carried out identifies in a better way the gesture of five fingers against the motion without extended fingers than the gesture of four fingers against that of.

4. Conclusions and Future Work

With the project's development, it can be concluded that both the built camera and the implemented library can support the detection and recognition of hand gestures. It should be noted that the technologies developed are relatively inexpensive. Based on this, the performance of this proposal is evaluated in a variety of circumstances, including illumination changes or camera movement, various sorts of gestures, and people's sensitivity or prejudice, among others. We will look at how infrared camera pictures may be utilized to segment, categorize, and recognise one-handed movements in various illumination circumstances in this post, which could be part of the consideration when compared with some of the works previously carried out and analysed in the related works section. However, the results obtained open the possibility of improving the technologies developed after reviewing the negative aspects that were found. One of them, for example, is the affectation of the identification of gestures due to the distance between the camera and the hand that makes the gestures, since when training vector support machines with a characteristic such as height and width, the samples used in training may not have been

sufficient to avoid being susceptible to errors due to such distance. Additionally, regarding the modified hardware, it is proposed to change the type of diodes used by another with a greater amplitude angle to cover more visible space or also to change the material used for filtering infrared light so that they adjust to the values used by LEDs. With the development of this work, it was possible to implement a biometric system consisting of a hardware module for the acquisition of infrared images and a software module for digital image processing and pattern recognition, capable of carrying out the tasks of capture, registration, and validation of the authenticity of people using the patterns of the vascular network of the dorsal aspect of the hand. Finally, it is possible to modify the library developed to avoid considering a part of the arm as part of the hand, which could also be the reason why there were some problems in identifying the size of the hand. In addition, the proposed method guarantees the extraction of the region of interest in a similar area for each of the images, avoids the use of fixation devices to centre the hand in the desired position, and eliminates the influences caused by small displacements and rotations that may be presented in the capture of images.

Data Availability

The data underlying the results presented in the study are available within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

References

- [1] P. Dhamanskar, A. C. Poojari, H. S. Sarwade, and R. R. D'silva, "Human computer interaction using hand gestures and voice," in *Proceedings of the 2019 International Conference on Advances in Computing, Communication and Control (ICAC3)*, pp. 1–6, Mumbai, India, December 2019.
- [2] V. Harini, V. Prahelika, I. Sneka, and P. Adlene Ebenezer, "Hand gesture recognition using OpenCV and Python," *New Trends in Computational Vision and Bio-inspired Computing*, vol. 174, pp. 1711–1719, 2020.
- [3] A. H. Hameed, E. A. Mousa, and A. A. Hamad, "Upper limit superior and lower limit inferior of soft sequences," *International Journal of Engineering & Technology*, vol. 7, no. 4.7, pp. 306–310, 2018, p.
- [4] S. Sengan, O. I. Khalaf, S. Priyadarsini, D. K. Sharma, K. Amarendra, and A. A. Hamad, "Smart healthcare security device on medical IoT using raspberry Pi," *International Journal of Reliable and Quality E-Healthcare*, vol. 11, no. 3, pp. 1–11, 2022, p.
- [5] S.-S. Ito, M. Ito, and M. Fukumi, "Japanese sign language classification is based on gathered images and neural networks," *International Journal of Advances in Intelligent Informatics*, vol. 5, p. 243, 2019.
- [6] C. Keskin, F. Kirac, Y. E. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors," in *Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 1228–1234, Barcelona, Spain, November 2011.
- [7] S.. Michahial, "Hand gesture recognition using support vector machine," *International Journal of Engineering Science*, vol. 4, p. 42, 2015.
- [8] S. Saleh, M. Rahman, and K. Pavel, "Comparative study on the software methodologies for effective software development," *International Journal of Scientific Engineering and Research*, vol. 8, pp. 1018–1025, 2017.
- [9] M. L. Thivagar, A. S. Al-Obeidi, B. Tamilarasan, and A. A. Hamad, "Dynamic analysis and projective synchronization of a new 4D system," in *IoT and Analytics for Sensor Networks*, p. Springer, Singapore, 2022.
- [10] J. S. Supancic, G. Rogez, Y. Yang, J. Shotton, and D. Ramanan, "Depth-based hand pose estimation: data, methods, and challenges," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, Washington, DC, USA, December 2015.
- [11] N. Umadevi and I.R Divyasree, "Development of an efficient hand gesture recognition system for human computer interaction," *International Journal of Engineering and Computer Science*, vol. 5, no. 2, 2016.
- [12] C. Wang and Z. Zhang, "Development CMOS sensor image acquisition algorithm based on wireless sensor networks," *Sensors and Transducers*, vol. 156, pp. 10–17, 2013.
- [13] S. Sengan, O. I. Khalaf, D. K. Vidya Sagar, D. K. Sharma, L Arokia Jesu Prabhu, and A. A. Hamad, "Secured and privacy-based IDS for healthcare systems on E-medical data using machine learning approach," *International Journal of Reliable and Quality E-Healthcare*, vol. 11, no. 3, pp. 1–11, 2022, p.
- [14] S. Jha, S. Ahmad, H. A. Abdeljaber, A. A. Hamad, and M. B. Alazzam, "A post COVID machine learning approach in teaching and learning methodology to alleviate drawbacks of the e-whiteboards," *Journal of Applied Science and Engineering*, vol. 25, no. 2, pp. 285–294, 2021, p.
- [15] Z. Ren, J. JMeng, and J. Yuan, "Depth camera based hand gesture recognition and its applications in Human-Computer-Interaction," in *Proceedings of the ICICS 2011 8th International Conference on Information, Communications and Signal Processing*, Singapore, December 2011.