



Research Article

Hand Gestures for Elderly Care Using a Microsoft Kinect

Munir Oudah¹, Ali Al-Naji^{1,2}, Javaan Chahl^{2,3}¹Electrical Engineering Technical College, Middle Technical University, Baghdad, Iraq.²School of Engineering, University of South Australia, Mawson Lakes SA 5095, Australia.³Joint and Operations Analysis Division, Defence Science and Technology Group, Melbourne VIC 3207, Australia.

✉ Corresponding authors. E-mail: ali.al-naji@unisa.edu.au

Received: Mar. 26, 2020; **Accepted:** Jun. 20, 2020; **Published:** Jul. 27, 2020**Citation:** Munir Oudah, Ali Al-Naji, and Javaan Chahl, Hand Gestures for Elderly Care Using a Microsoft Kinect. *Nano Biomed. Eng.*, 2020, 12(3): 197-204.**DOI:** 10.5101/nbe.v12i3.p197-204.

Abstract

The link between humans and computers, called human-computer interaction (HCI) techniques, has the potential to improve quality of life, where analysis of the information collected from humans through computers allows personal patient requirements to be achieved. Among them, a computer vision system for helping elderly patients currently attracts a large amount of research interest to avail of personal requirements. This paper proposes a real-time computer vision system to recognize hand gestures for elderly patients who are disabled or unable to translate their orders or feelings into words. The proposed system uses a Microsoft Kinect v2 Sensor, installed in front of the elderly patient, to recognize hand signs that correspond to a specific request and sends their meanings to the care provider or family member through a microcontroller and global system for mobile communications (GSM). The experimental results illustrated the effectiveness of the proposed system, which showed promising results with several hand signs, whilst being reliable, safe, comfortable, and cost-effective.

Keywords: Real-time; Non-contact computer vision system; Microsoft kinect sensor; Arduino nano microcontroller; Global system for mobile communication (GSM); Elderly care

Introduction

The aged population in the world is increasing by 9 million per year and it is expected to reach to more than 800 million people by 2025 [1]. The increase in the number of elderly people requires many healthcare organizations and homecare services. Although homecare services enable elderly patients to remain at home and thus reduce the effective costs resulting from long hospital stays and minimizing visits to emergency departments, it is hard for the elderly to communicate with family membership or care provider when they need help if they request something such as

a meal, water, toilet and other personal requirements [2]. Human-computer interaction (HCI) based on a computer vision system using hand gesture may provide an important homecare service and help elderly patients [3] who are deaf-mute or have lost speech ability because of stroke, blindness or hemiplegia which make them unable to convey or translate their sensation to the care provider easily and quickly.

The current glove attached sensors for recognizing hand gestures, such as flex sensors [4], gyroscope and accelerometer sensors [5] are restrictive and uncomfortable for patients. In addition, they may cause skin damage to people who suffer from burns and

those with sensitive skin. To address these challenges, a non-contact sensor based on a camera was used in this study to recognize the elderly hand signs without any physical contact. Moreover, the obtained hand sign corresponding to a specific request was reported to connect the care provider through a global system for mobile communication (GSM) modem.

This system used a real-time computer vision system based on a Kinect v2 Sensor to recognize hand and provided natural interaction without using direct-attaching sensor which may reduce the communication gap between patients and care providers or other family members.

Related works

The hand gesture recognition systems can be classified into two approaches. The first approach is the contact approach using wearable gloves with a direct-attached sensor to provide physical response depending on the type of sensors such as flex sensors [4], gyroscope and accelerometer sensors [5], tactiles [6], and optic fibers [7]. All these sensors are used to give the output response related to capturing hand movement and position by detecting the proper coordination of the site of the forearm and fingers. However, sensor-based glove approach needs to be attached to the computer physically [8], which may obstruct the facility of interaction between patients and computers. In addition, some of these sensors are expensive and may cause sensor fault in the case of power failure [9].

The second approach is the non-contact approach mainly utilizing a computer vision system based on a camera such as thermal camera [10], RGB camera, depth camera, time of flight camera (TOF) [11], and night vision camera [12]. Many image and video processing algorithms are used with this approach to capture images or videos [13, 14]. These algorithms are dealing with a region of interest (ROI) through pixel processing and extract features that help to track and clarify hand sign. Therefore, this non-contact approach provides natural interaction where there is no need to wear any external hardware. However, some disadvantages may arise with variation in illumination levels, orientations and positions of the subject in the captured video. Therefore, several pre-processing operations need to be implemented to address these issues. Several methods were used in improving the algorithm at different mechanism achievements for detecting and tracking the static or dynamic hand

sign such as Haar-like features [15], convex hull mechanism [16], contour matching technique [17], and skin color method [18]. However, these methods have some drawbacks if they are compared with the first approach such as the Haar technique requires an accurate type of data for learning and training to detect each hand. Also, the Convex hull method and contour matching technique are complete with the convexity defect realize from a chamber between each finger. To be able to distinguish the hand signs, the fingers should be completely visible with the area between each of them, while skin color has issues consisting of lighting condition variations and problems when the background has the same color with the skin. On the other hand, in the recognition procedure, several motivating mechanisms are applied inclusively rule-based [19], dynamic time warping (DTW) [20], hidden Markov Model (HMM) and support vector machine (SVM) [21]. However, these recognition methods may sometimes use data set, which increases accuracy over the rule-based method, getting a valid outcome which is considered simple and suitable for a small amount of data. Finally, these approaches provide natural interaction with cost-effectiveness. In this study, we developed a real-time non-contact system based on a Microsoft Kinect Sensor, Arduino Nano microcontroller and GSM to capture hand signs of elderly patients and send their request to the care provider accurately and as fast as possible.

Experimental

Participants and experimental setup

The paper was conducted using 4 elderly participants (3 male and 1 female) between the age of 60 and 75 years old and 1 adult (33 years old). This study adhered to the Declaration of Helsinki ethical principles (Finland 1964) where a written informed consent form was obtained from all participants after a full explanation of the experimental procedures. The experiment was for approximately 1-h for each participant at home environment and repeated at various times to obtain sufficient outcomes. The Microsoft Kinect v2 sensor was installed at an angle of 0° and a distance of approximately 1-1.5 m with a resolution of 1920 × 1080 and a frame rate of 30 fps. The Kinect sensor was connected to a laptop with a conversion power adaptor and a standard development kit (Microsoft Kinect for Windows SDK 2.0) installed. Fig. 1 shows the experimental setup of the proposed system.

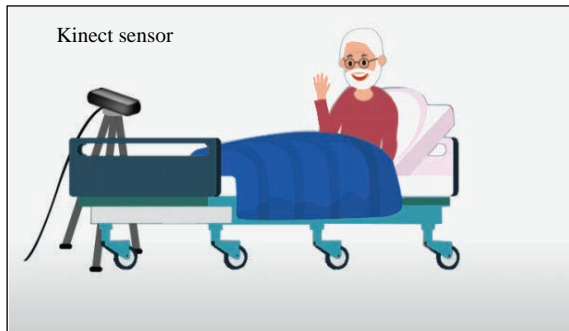


Fig. 1 Experimental setup of the proposed system.

System design

The schematic diagram of the proposed system is shown in Fig. 2. The system design of the proposed system can be divided into 3 main parts: the Microsoft Kinect sensor, Arduino Nano microcontroller and GSM Module Sim8001.

Microsoft Kinect sensor

Microsoft released the Kinect v1 as a peripheral device in 2010; the Microsoft company released XBOX360 for game purposes. With the increasing request in the store, the company modified it to be appropriate for operating systems like Windows using Microsoft's Kinect Standard Development Kit (Microsoft Kinect for Windows SDK 2.0) and power converter [22-25]. With this, it made it possible for developers and researchers to do development on a computer. After that, the new release in 2014 considered the subsequent generation of the sensor (Kinect v2) which improved rendering, precision and wider field of view [22-25]. This is because the v2 release utilized a ToF technology [26] instead of light coding technology [27] utilized in Kinect sensor v1. Differentiation between the two releases (Kinect v1, v2) technology was explained in previous studies [25-28]. Fig. 3 shows the outer view of the Microsoft Kinect sensor v2 for Xbox One.

Kinect sensor v2 includes 3 visual sensors, a RGB camera, an IR sensor, an IR projector that provides outputs, an RGB image, and a depth image. These

features permit body tracking, 3D human rebuilding, human skeletal tracking, and human joint tracking. Because Kinect v2 is supplied with PC adapter and has particular advantages at a low cost, and is intended for the gaming uses, it has been a common equipment for many biomedical implementations in both research and clinical applications [29].

Arduino Nano microcontroller

The Microcontroller, based on Arduino, type Nano, acts as an interface between GSM module and computer, which links with the module through 2 digital pins and with a computer through USB serial. This microcontroller has 14 digital I/O pins and 8 analog pins. Also, the clock frequency is 16 MHz [30]. It receives data from Matlab and processes it to control message sending. In addition, it has some advantages such as small size, cheap, and easy to program with an open-source platform.

GSM module SIM8001

The GSM (module SIM8001) is a small modem approximately 0.025 m² that can operate in a voltage range from 3.4 to 4.4 V [31], which is used for communication purposes, such as sending and receiving SMS messages, GPRS data and making voice calls. Therefore, it is suitable for usage in sending a patient request to the care provider cellphone containing 5 messages controlled by data process in a microcontroller depending on a Matlab program command. Therefore, the GSM module transmitter and receiver (TX, RX) is connected with 2 digital pins of a microcontroller. In addition, the Vcc, ground pins of the module are connected with 5 V charger through the DC to DC step down buck converter LM2596 to avoid the drop voltage of the microcontroller.

Image and video algorithms

This section characterizes the proposed camera module architecture based on Kinect output as shown in Fig. 4. As the first step, the Kinect RGB camera sensor was used for acquiring an RGB color frame

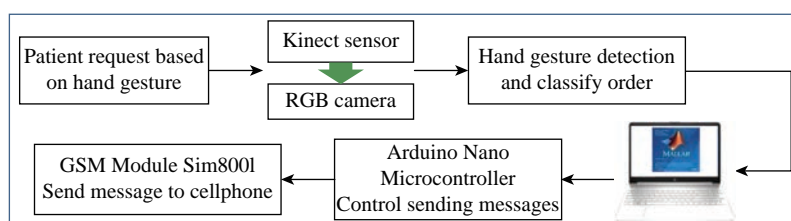


Fig. 2 The schematic diagram of the proposed system.



Fig. 3 Microsoft Kinect sensor v2 for Xbox One.

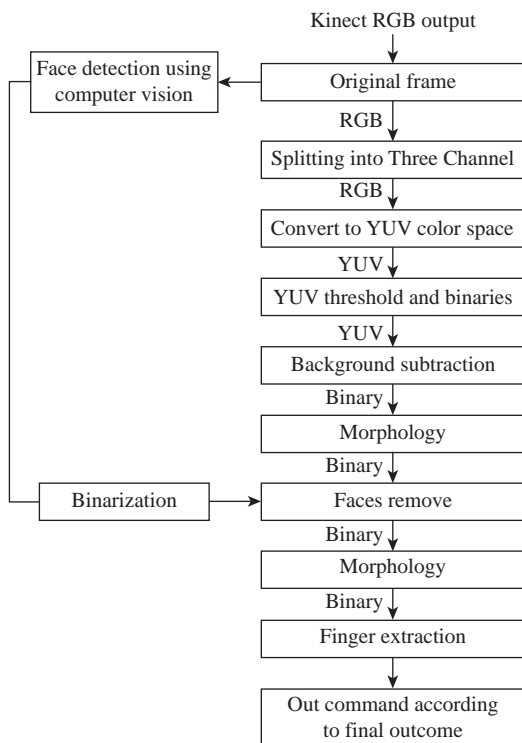


Fig. 4 Camera module architecture based on Kinect sensor.

continuously, which provided resolution 1080X1920 for 30 frames per second (fps) and setting for a single frame per trigger. After that, the algorithm based on the computer vision system was used to detect the patient's face before processing the frame. Thereafter, apply the image processing algorithm to the image frame's in order to process and convert RGB color space to YUV color space [32]. After getting YUV channels, the threshold was applied to these channels to segment skin color tone. The next step was to use morphology

operations such as image erosion, dilation and filtering to detect and track the patient's hand.

The following points clarify the function of each block:

(1) The Kinect RGB-camera was used as acquisition device to acquire video frame in real-time.

(2) The first step used KLT algorithm to detect face in a live video then handled the result to use it in another step.

(3) Then the RGB video frames split into sequence of frames in order to be processed with image processing in the next steps.

(4) Convert RGB colour space into YUV colour space splitting into 3 channels.

(5) Apply thresholding to YUV channels to extract skin colour.

(6) The background was subtracted by the previous step and cleaned using morphological operations in this step in order to smooth the edge of skin detected and remove small objects.

(7) User morphology operation to enhance the result and segmented the face only by subtracting the detected face by the first step with the previous result.

(8) Morphological operations were used to enhance results from the binary image and extract the hand palm in order to count the number of fingers using Connected components function in a binary image where each number of finger appeared related to the specific request.

The morphological operations were applied to the logical image which had 2 values, 0 and 1. Therefore, it can easily deal with the ROI such as object extraction, image subtraction, image masking, elimination of image noise and enhancement of the outcome result.

As an image resulted from the preceded operation, it can be filtered with erosion; to eliminate undesired pixels, while pixel of hand palm was set to be 1, then every pixel near the border would be erased count on the radius of the palm centre. 5 hand fingers could be found by subtracting the resulting image frame from the preceded frame. The final step was to detect and count white areas of the finger appearance that represented and identified the order. Each hand sign was introduced as a request, involving 1 to 5 finger signs where they signaled "water," "meal", "toilet", "help" and "medicine".

Result and Discussion

In this experiment, the 5 hand signs were discovered and familiarized by using the suggested method in real-time. The Kinect sensor captured 30 frames per second (fps) and was used in the tested method with a computer Elite-book Intel(R) Core(TM) i7-2620M CPU@(2.70 GHz) 64-bit, Windows 10 operating system and Matlab R2018a program. The empirical outcomes offered 4 overtures including a mainframe, hand tracking, hand segment and finger extraction. The hand segmentation method based on skin colour using RGB Kinect camera was followed by morphological processing to minimize noise and filling punctures into the hand surface, which was then represented as binary form black (as background elimination) and white region (hand).

Finally, the fingers were extracted by doing some of the morphological processes and was utilized to give the hand sign and identify the request of the hand signs according to finger counting. Using this execution, it is possible to identify the request of each hand sign accurately identical to the request as specified by the elderly participants and the care provider at the preceded procedure. Fig. 6 shows the result of the proposed system.

The system presented 5 hand gestures (1, 2, 3, 4 and 5) using a finger counting technique. Our analysis with regard to all the recorded tested samples with results are presented in Table 1 and Fig. 5.

The system was tested at 1 h for each participant, where the total number of tested gestures could be calculated as follows:

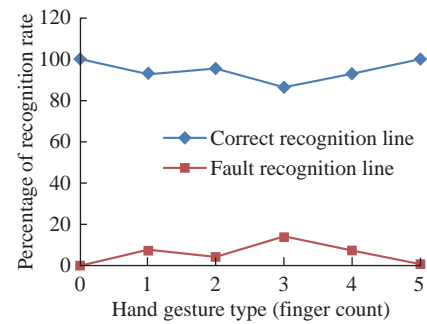


Fig. 5 The recognition rate of the total tested gestures.

$$T = P \cdot I \cdot G \cdot M \tag{1}$$

where T is the total number of tested per gesture, P is the number of participants, I is the number of repetition test at a different time per participant, G is the number of type gesture tested, and M is the repetition test at 1 h for every gesture. Using this equation, we can easily calculate the total of the tested gestures. In this case, the 32 samples for every type of gesture was tested for elderly participants and 192 presented total number of gesture tested for elderly participants. Add the test result for the adult person where M equal to 4 and I equal to 3. The test gave 12 samples for every type of gesture and presented 72 total number of gestures tested for the adult participants.

As mentioned above, the total number of test per unit gesture was 44 for all participants and 264 tested samples for overall tests. The result presented for false recognition was because detection was more sensitive for gesture number “3” than the other gestures. To overcome this problem, we processed the number of frames such as 5 frames, picked out the last result, and so on, which might be slightly time-consuming but decreased the fault rate. All the requests in Fig. 6 are

Table 1 The outcome analysis of the total number of tested gestures per participant

Hand gesture type	Total number of samples per tested gestures	Number of recognizable gestures	Number of Un-recognizable gestures	Percentage of correct recognition for total number of sample gestures (%)	Percentage of false recognition for total number of sample gestures (%)
0	44	44	0	100	0
1	44	41	3	93.18	6.82
2	44	42	2	95.45	4.55
3	44	38	6	86.36	13.64
4	44	41	3	93.18	6.82
5	44	44	0	100	0
Total	264	250	14	--	--

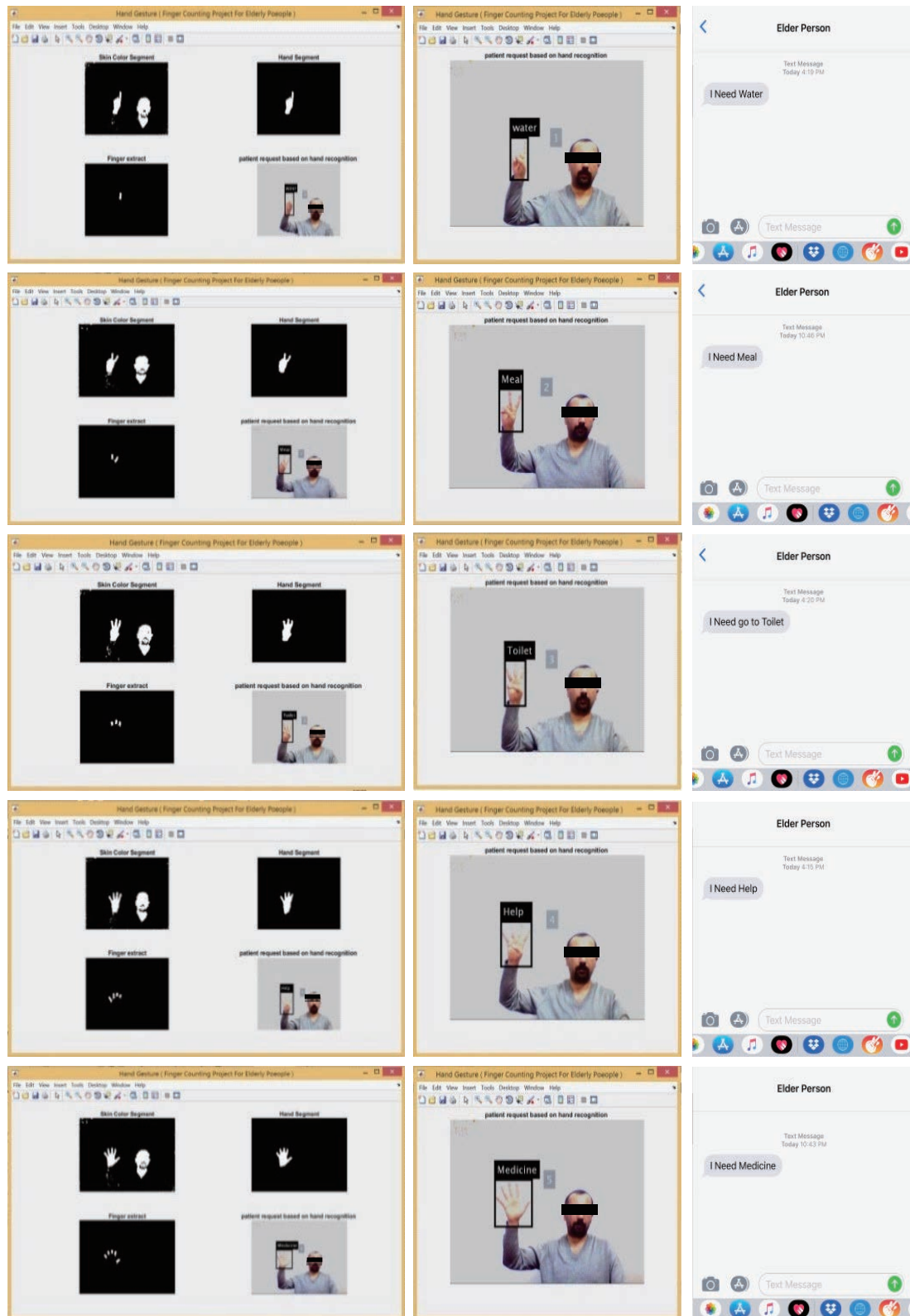


Fig. 6 The obtained results of the proposed system.

related to the number of finger appearance.

Although the proposed system was accurate, cost-effective, easy to use and portable, it has some limitations. The first limitation is that some older patients may have difficulties remembering the signs

or gestures. In addition, some medical events such as hand convulsionse and Parkinson's disease, might impact the stabilization and motion of the hand [33]. Therefore, a simple approach should be used with regard to these cases. The second limitation is that the

proposed system may have some background issues similar to skin color tone and low light condition.

Conclusions

In this paper, a real-time computer vision system to recognize hand gestures for elderly patients was proposed using Microsoft Kinect sensor v2, Arduino Nano Microcontroller and GSM. The 5 hand signs recognised included the open palm gestures, 1, 2, 3 and 4 sign fingers which identified the elderly patients' requests such as water, meal, toilet, help and medicine, sent as a message through the GSM Module Sim800l controlled by an Arduino Nano Microcontroller. The proposed system was implemented under daylight conditions and offered the elderly people an easy and simple way to communicate with the family member or care provider. Although the proposed system was accurate, cost-effective, easy to use and portable, further studies on robustness, low light condition and detection range are needed to achieve better outcomes.

Conflict of Interest

The authors declare that no competing interest exists.

References

- [1] T. Truelsen, R. Bonita, and K. Jamrozik, Surveillance of stroke: A global perspective. *Int. J. Epidemiol.*, 2001, 30(suppl_1): S11.
- [2] A. Chaudhary, J.L. Raheja, A health monitoring system for elder and sick persons. *International Journal of Computer Theory and Engineering*, 2013, 5(3): 428-431.
- [3] W. Chen, Gesture-based applications for elderly people. *Proceedings of International Conference on Human-Computer Interaction*, 2013: 186-195.
- [4] V. Sharma, V. Kumar, S.C. Masaguppi, et al., Virtual talk for deaf, mute, blind and normal humans. *Proceedings of 2013 Texas Instruments India Educators' Conference*, 2013: 316-320.
- [5] H. Basanta, Y.P. Huang, and T.T. Lee, Assistive design for elderly living ambient using voice and gesture recognition system. *Proceedings of 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017: 840-845.
- [6] A. Rajamohan, R. Hemavathy, and M. Dhanalakshmi, Deaf-mute communication interpreter. *Int. J. Sci. Eng. Technol.*, 2013, 2(5): 336-341.
- [7] S.A. Mehdi, Y.N. Khan, Sign language recognition using sensor gloves. *Proceedings of the 9th International Conference on Neural Information Processing*, 2002, 5: 2204-2206.
- [8] N.A. Ibraheem, R.Z. Khan, Survey on various gesture recognition technologies and techniques. *Int. J. Comput. Appl.*, 50(7): 2012.
- [9] K.A. Sunitha, P.A. Saraswathi, M. Aarthi, et al., Deaf-mute communication interpreter-a review. *Int. J. Appl. Eng. Res.*, 2016, 11(1): 290-296.
- [10] J. Appenrodt, A. Al-Hamadi, and B. Michaelis, Data gathering for gesture recognition systems based on single color, stereo color-and thermal cameras. *Int. J. Signal Process. Image Process. Pattern Recognit.*, 2010, 3(1): 37-50.
- [11] S. Hussmann, T. Liepert, Robot vision system based on a 3D-ToF camera. *Proceedings of 2007 IEEE Instrumentation & Measurement Technology Conference*, 2007: 1-5.
- [12] W.F. Abaya, J. Basa, M. Sy, et al., Low cost smart security camera with night vision capability using Raspberry Pi and OpenCV. *Proceedings of 2014 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management*, 2014: 1-6.
- [13] A.M. Tekalp, *Digital video processing*. Prentice Hall Press, 2015.
- [14] A. McAndrew, An introduction to digital image processing with matlab notes for SCM2511 image processing. *Sch. Comput. Sci. Math. Victoria Univ. Technol.*, 2004, 264(1): 1-264.
- [15] C.C. Hsieh, D.H. Liou, Novel Haar features for real-time hand gesture recognition using SVM. *J. Real-Time Image Process.*, 2015, 10(2): 357-370.
- [16] P. Barros, N.T. Maciel-Junior, B.J.T. Fernandes, et al., A dynamic gesture recognition and prediction system using the convexity approach. *Comput. Vis. Image Underst.*, 2017, 155: 139-149.
- [17] Y. Pang, L. Ye, X. Li, et al., Incremental learning with saliency map for moving object detection. *IEEE Trans. Circuits Syst. Video Technol.*, 2016, 28(3): 640-651.
- [18] W. Wang, J. Pan, Hand segmentation using skin color and background information. *Proceedings of 2012 International Conference on Machine Learning and Cybernetics*, 2012, 4: 1487-1492.
- [19] T. Hachaj, M.R. Ogiela, Rule-based approach to recognizing human body poses and gestures in real time. *Multimed. Syst.*, 2014, 20(1): 81-99.
- [20] S. Axelrod, B. Maison, Combination of hidden Markov models with dynamic time warping for speech recognition. *Proceedings of 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2004, 1: I-173.
- [21] M.E. Mavroforakis, S. Theodoridis, Support vector machine (SVM) classification through geometry. *Proceedings of 2005 13th European Signal Processing Conference*, 2005: 1-4.
- [22] M. Samir, E. Golkar, and A.A.A. Rahni, Comparison between the KinectTM V1 and KinectTM V2 for respiratory motion tracking. *Proceedings of 2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 2015: 150-155.
- [23] C. Kim, S. Yun, S.W. Jung, et al., Color and depth image correspondence for Kinect v2. *Advanced multimedia and ubiquitous engineering*, Springer, 2015: 111-116.
- [24] L. Yang, L. Zhang, H. Dong, et al., Evaluating and improving the depth accuracy of Kinect for Windows v2. *IEEE Sens. J.*, 2015, 15(8): 4275-4285.
- [25] H. Sarbolandi, D. Lefloch, and A. Kolb, Kinect range sensing: Structured-light versus Time-of-Flight Kinect. *Comput. Vis. image Underst.*, 2015, 139: 1-20.
- [26] C.D. Mutto, P. Zanuttigh, and G.M. Cortelazzo, *Time-of-flight cameras and microsoft kinect (TM)*. Springer Publishing Company, Incorporated, 2012.
- [27] A. Al-Naji, K. Gibson, S.-H. Lee, et al., Real time apnoea monitoring of children using the Microsoft Kinect sensor: A pilot study. *Sensors*, 2017, 17(2): 286.

- [28] A. Al-Naji, J. Chahl, Detection of cardiopulmonary activity and related abnormal events using microsoft kinect sensor. *Sensors*, 2018, 18(3): 920.
- [29] Y. Li, Hand gesture recognition using Kinect. Proceedings of 2012 IEEE International Conference on Computer Science and Automation Engineering, 2012: 196-199.
- [30] R.H. Kumar, A.U. Roopa, and D.P. Sathiya, Arduino ATMEGA-328 microcontroller. *Int. J. Innov. Res. Electr. Electron. Instrum. Control Eng.*, 2015, 3(4): 27-29.
- [31] S. Mluyati, S. Sadi, Internet of Things (IoT) Pada Prototipe Pendeteksi Kebocoran Gas Berbasis MQ-2 Dan SIM800L. *J. Tek.*, 2019, 7(2): 64-72.
- [32] T. Nguyen-Trang, A New Efficient Approach to Detect Skin in Color Image Using Bayesian Classifier and Connected Component Algorithm. *Math. Probl. Eng.*, 2018.
- [33] T. Pringsheim, N. Jette, A. Frolkis, et al., The prevalence of Parkinson's disease: A systematic review and meta-analysis. *Mov. Disord.*, 2014, 29(13): 1583-1590.

Copyright© Munir Oudah, Ali Al-Naji, and Javaan Chahl. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.