Vision-Based Control By Hand-Directional Gestures Converting To Voice

Fahad Layth Malallah, Khaled N. Yasen Mustafa Gh. Saeed, Sadeer Dheyaa Abdulameer, Ammar Waysi Altuhafi

Abstract: Currently, remote vision-based control, which is converted into a voice, is a research becoming one of the major focuses among researchers due to various applications in terms of Human-Computer-Interaction (HCI). The proposed idea is to learn a robot how to receive hand directional gestures as left, right, up and down, and pronounce these directions according to the originated directions. To implement this, a technique is proposed by using index-finger of a hand in order to point out random directions, in which the feature extraction is based on the area of the object pixels. It is static hand directional gesture recognition. This method starts by inserting a video to extract frames, which are processed one by one to output 4-possible-direction. After that, each specific direction will switch on its corresponding voice pronouncing the direction sign as a robot speaking. Experiments were conducted on 10-video frames for testing by using 10 individuals. The total frames for this experiment is having 442 frames. Results demonstrate high successful recognition rate in terms the performance accuracy up to 94.57%.

Keywords: Hand Gesture to Voice, Pattern Recognition, Feature Extraction, Expert System, Computer Vision, Data Science.

1. INTRODUCTION

In past decades, development of information technology imposed needs for new types of human-computer-interaction by using the embedded systems for better and more comfortable lives [1]. One of these development is hand gesture recognition [2], which is considered as behavioral type of biometric [3, 4], which has become more and more popular in applications like intelligent sensing, robot control, smart guidance. Its applications range from sign language recognition, control system remotely in Robotics, through medical rehabilitation, authentication, also in 3D control where you can move. However, one of the major challenges in vision-based hand gesture recognition is to recognize the hand gestures effectively in different background conditions. Background may vary from place to another depending on the environment conditions, especially, when background conditions change due to varying illumination conditions, occlusion, dynamic or moving objects in background, cluttered or distorted objects in background scene etc. The systems must have fast response times and high recognition accuracy, as well as be easy to learn and must provide a high degree of user satisfaction, in which the real-time recognition system, these conditions should be taken into consideration as these challenges present in real time scenario, which affects the robustness of the system in recognizing the hand gestures [5].

- *Fahad Layth Malallah, Khaled N. Yasen Mustafa Gh. Saeed, Sadeer Dheyaa Abdulameer, Ammar Waysi Altuhafi*
- *Cihan-Slemani University, Computer Science, Slemani, Iraq.*
- *Departement of Computer Science, Cihan-University, Erbil, Iraq*
- *Fahad.layth@sulicihan.edu.krd, khalid_aldabbagh@cihanuniversity.edu.iq, mostafa.swe@gmail.com, sadeer.alatter@gmail.com, ammar_wysi@yahoo.com*

The scope of this research is as following, it is fixed background of the individual to gain fast processing, as well as it is considered as indoor signing operation, to be invariant to the weather changing. This paper aims to introduce an efficient algorithm to teach robot regarding receiving several control signals such as hand-direction-gesture captured as up, down, left, and right by using forefinger (index-finger) of the hand. Furthermore, teach the computer robot to pronounce these directions accordingly with a high performance. Besides that, the aim of this paper is trying to solve touch-based control problems to avoid any kind of symptoms that cause allergic to other unaffected people. In other words, the paper opens the door to the safe and healthy systems against any kind of touch infection, especially, if such these devices are in general usage. The contribution knowledge of this paper is using index-finger of a hand to point out to random different directions to be considered as command to control any required device remotely by just using a the proposed algorithm embedded to smart camera. The organization of the paper is as follows; Section 2 covers literature review regarding overall hand gesture previous works and types, Section 3 explains the proposed framework and the concept methodology comprises the required features and their classifications. Section 4 presents the experiment details of this research, while Section 5 presents the results and discussions. Finally, Section 6 concludes this research and tailing with a possible future work.

2. LITERATURE REVIEW

Control using vision-based via hand gesture is considered an application of Hand Gesture Recognition which is available in the literature that can be broadly divided into two types, Touch-Based and Contactless-Based hand gesture. The former is described as giving multi-hand gesture by touching the human palm with a sensing scanning device [6], example of such work is human multi-touch hand gesture authentication [7], while the latter one can be characterized as hand gesture signals are transmitted to a computer for recognition operation remotely without any touched device [8-10]. Contactless-based has also two branches. Firstly, handheld device-based hand gesture which employs sensors (mechanical or optical) attached to a human hand that transduces hand waving signals for determining the hand

posture or path trajectories. In this method, the user has to carry a load of cables which are connected to the computer and hinders the ease and naturalness of the user interaction, e.g., glove based hand gesture [1]. Secondly, vision-based hand gesture, which is described as a computer understanding the hand gesture through camera or sensors [11], e.g., air-writing characters can be recognized similar to motion gestures in free space by hand [12]. Recently, in 2018, Hand Gesture has been used to control Drones such as drone movement, such as takeoff, landing, hovering and so forth. A set of five gestures are studies [13]. Also in [14], another work for Web-based interactive drone control using hand gesture is presented in 2018. Furthermore, it is using for controlling the desktop cursor as in [15]. Another interesting work presented in 2018 as well, Smart infrastructure has become an attract research by using Finger Angle-Based Hand Gesture Recognition for Smart Infrastructure Using Wearable Wrist-Worn Camera, the recognition here is based on finger segmentation and template matching as in[16]. Therefore, the challenge is still open in this framework by improving the recognition rate and the robustness of the hand segmentation. In this paper, a new method of hand gesture recognition for vision-based control will be presented based on hand direction of the fingertips that are going to be used in different application of remote control. The advantage of this method is accurate in terms of recognition rate.

3. Framework Design

The implementation of converting vision-based hand gestures to its corresponding voice is explained in this section. The operation is started by capturing video by a camera, then frame by frame is captured and processed separately. Then, the stage of hand detection is based on hand color filtering after converting from RGB to YCbCr color space, which is more robust against illumination variance. Next, the hand object is segmentation based on black and white conversion with a predefine threshold to be in a form of a binary image as only hand, in order to do classification of the direction, the obtained object is divided into four directions (signals), i.e., left, right, up and down directions detailed in next section. The framework design of this methodology is depicted in Figure 1. Once the directions are identified, the automatic voice is switched on based on the hand gesture.

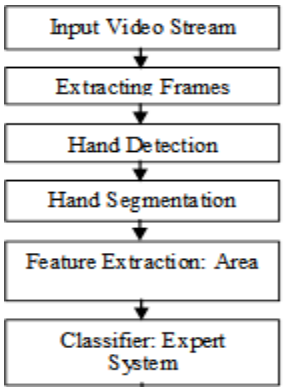


Fig. 1. *Diagram depicts the overall proposed methodology.*

The operation starts by detecting the hand as shown in Figure 2-1 and Figure 2-2 by using hand color as YCbCr [17]. After that, converting operation to the binary image is applied to segment the region of interest (ROI) as shown in Figure 2-3.

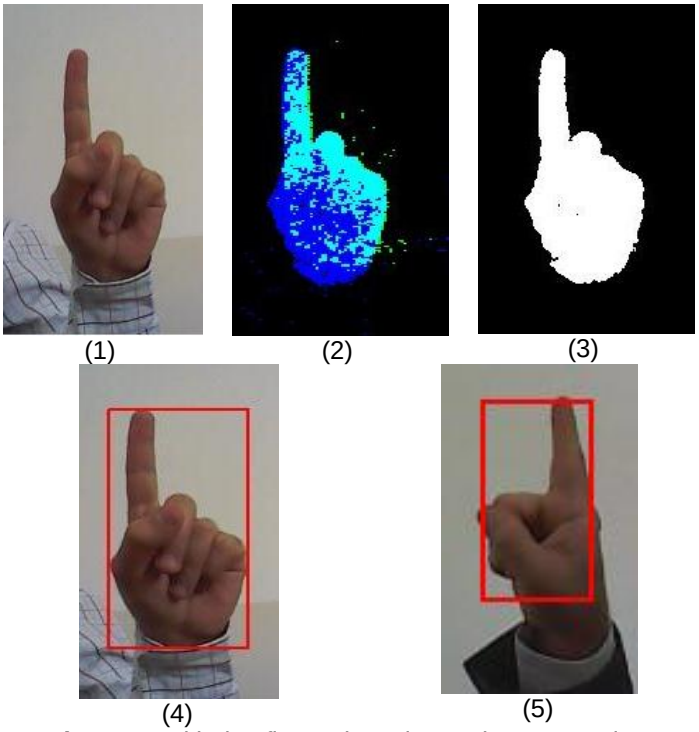


Fig. 2. *Hand index-finger detection and segmentation illustration.*

It is a good idea to apply some image processing tools such as filtering noise in order to avoid converting noises captured by the camera to be part of the segmented object, and then searching for the largest object in order to ensure removing all objects except the hand object as in Figure 2-4, in which this image has a tracker border around the target hand, also another example of hand ROI is depicted in Figure 2-5. Now, the feature vector will be area of the hand object (convex), to be passed to the classifier, which is expert system.

3.1 Hand Localization

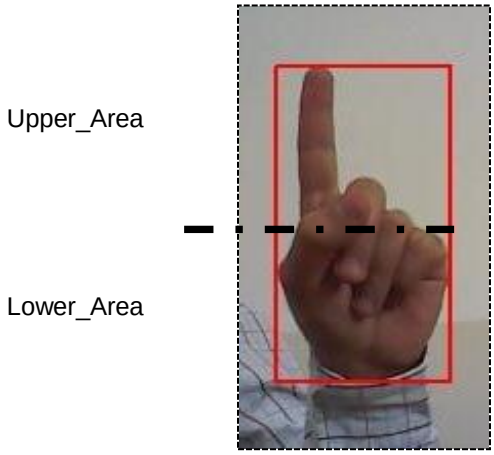
This operation starts by converting from RGB to YCbCr so as to split luminance from the chrominance. The blue and red chrominance are exploited to get the hand skin color (for this paper the color skin is modeled for white people). Also, hand detection operation is applied based on filtering Cb and Cr threshold [17] as following:

$$Cb \text{ limitation: } 77 < Cb < 120$$
$$Cr \text{ limitation: } 137 < Cr < 163$$

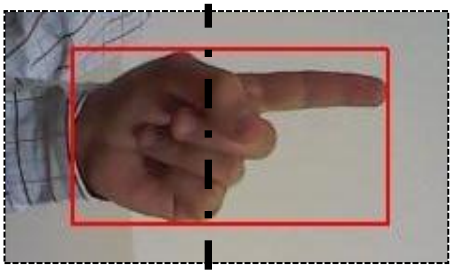
Then, the white and black operation is applied to the image. Afterward, median filter is applied for noise removing, which is replacing each pixel value in an image with the middle value of its neighbors after ordering them from the smallest to the highest pixel value. Then, extracting the bounding coordinates around the hand object as p1(x1,y1), p2(x2,y2), p3(x3,y3) and p4(x4,y4). According to these bounding points, a rectangular real-time tracker will be pictured around the ROI.

3.2 Feature Extraction

The main feature used in this proposed work is the area of the ROI object, which is the number of the hand pixels. This idea of feature extraction is adapted from the paper as in [2] published in 2017. The features are based on ROI pixel areas of the hand object. The final features are extracted as following: Upper_Area, Lower_Area, Left_Area and Right_Area as shown in Figure 3. These four features are deemed as the inference of the expert system classifier. In case the number of columns of the digital image hand is less than rows, then, the border (dotted line) will be assumed horizontally across the hand digital image, to separate the hand to either the Upper-Area or Lower_Area as shown in Figure 3-1. On the other hand, if columns are more than rows as shown in Figure 3-2, then the dotted line will be assumed vertically across the ROI, to get the Left-Area and Right_Area.



(1)



(2)

Fig. 3. *Shows different area features for the hand.*

In terms of pixel area calculation, it is defined as the summation of whatever pixel value has "1" throughout the ROI, area formula can be computed as the in Eq. (1):

$$Pixel_{Area} = \begin{cases} \sum_{y=1}^{M}\sum_{x=1}^{N} pixel(x,y) & , \quad pixel(x,y) = "1" \\ 0 & , \quad pixel(x,y) = "0" \end{cases} \quad (1)$$

Where (x,y) are pixel coordinate of the object. M and N are the ROI dimension.

3.3 Classification (Expert System)

Expert system [18] has been utilized as a classifier in order to output the result of each frame, the reason of selecting expert system is due to the nature of the feature extraction as area, which will be considered the condition of the judgment. Expert system classifier is based on if-else as a programming statement, which works by setting the knowledge bases that are described as set of rules as If-Then-Else as a programming language statement. Besides that, setting knowledge base, inference also must be entered to the expert system to be depended on it so as to enable these rules how to make the decision. Logically, inferences are represented by the extracted features. In other words, inference of the expert system in this paper is the area of the hand object such as pixel of the Upper_area, Lower_area, Left_area and Right_area.

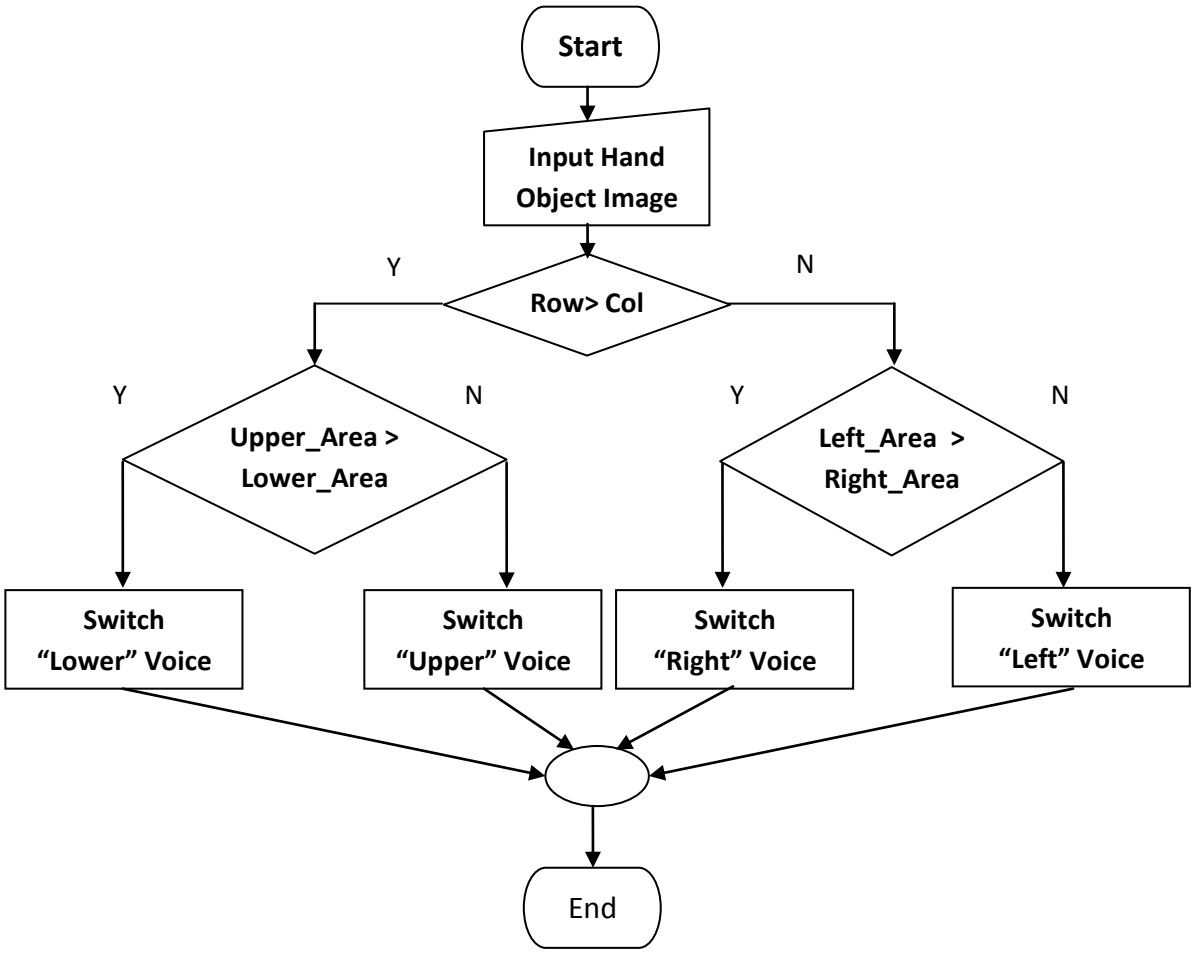


Fig. 4. *Depict the flowchart of the classifier how it is work for more explanation.*

The decision making for the four directions are described well in [2], which are:

1- Direction Left: if the Left_Area is less than the Right_Area as shown in Figure 3.
2- Direction Right: if the Left_Area is larger than the Right_Area as opposite of Figure 3.
3- Direction Up: if the Upper_Area is less than the Lower_Area as shown in Figure 3.
4- Direction Down: if the Upper_Area is larger than the Lower_Area as opposite of Figure 3.

Furthermore, the whole operation of the proposed paper is illustrated in the flowchart shown in Figure 4. The process starts when the ROI of the hand is detected and segmented thoroughly. Then, the first checking is occurred between the length of the ROI rows and columns, if the rows number larger than columns. Then, that means the predicted direction is either upper or lower. By keeping on checking, if the feature area Upper_Area is larger than Lower_Area, it means the direction sign is pointed to the lower. Otherwise, the direction is point to the upper. On the other hand, by going back to the first checking, if the number of columns is larger than rows of the ROI, then, that means the predicted direction is either left or right. By continuing checking, if the feature area Left_Area is larger than Right_Area, it means the direction sign is pointed to the right. Otherwise, the direction is point to the left. For more explanation, a mathematical modeling can be formed as in Eq. (2): If direction is referred to d, r ROI rows and c is ROI columns.

$$f(d) = \begin{cases} r > c \begin{cases} Upper_{Area} > Lower_{Area} & , \ d = upper \\ Upper_{Area} < Lower_{Area} & , \ d = lower \end{cases} \\ r < c \begin{cases} Right_{Area} > Left_{Area} & , \ d = left \\ Right_{Area} < Left_{Area} & , \ d = right \end{cases} \end{cases} \quad (2)$$

4. EXPERIMENT

In order to evaluate the performance of the proposed method, ten-individual have been captured to be used in the testing stage. These individuals asked to originate an control signal by their hand in front of a camera, and the signals are given by pointing out with their index-finger as pointing to a specific direction. As it is normal, the four directions learned to the computer are as follows: up, down, left, right either clockwise or anti-clockwise with any free random hand movement. About the characteristics of the taken video is as follows: 15 frames / sec. The time duration of each video is not similar as it is randomly taken ranging from 2 till 5 sec. The dimension as width is 320 and height is 240 with Bit/Pixel is 24. The experiment is run by doing training (as the proposed expert system based on the area inference) and testing. For the evaluation process is based on the predicted result, weather the originated hand direction recognized correctly or wrongly. Then the overall error is computed statistically. Finally, it is significant to mention the characteristics of used hardware,

which is used for capturing video recording; its brand is LOGITECH, as a VGA technique, 70Hz with 5MP resolution. As well as Matlab R2018a software used with windows 7 operating system installed in a personal computer having core2due, 4G-RAM memory and 2GHz CPU.

5. Result and Discussion

The total accuracy of the system is computed by subtracting the average error rate (E) from 100% as in Eq. (3):

$$\text{Accuracy}\% = 100\% - \text{Error} \quad (3)$$

In this research, E used for this testing is measured by dividing the number of wrongly recognized direction on the total number of the originated direction. As formulated in Eq. (4).

$$E\% = \frac{\text{No.of Wrongly Predicted Directions}}{\text{No.of total Frames}} \times 100\% \quad (4)$$

Results of the 10 individuals are reported in Table 1, and mentioned number of the wrongly as well as correctly predicted frames.

Table 1. *Results of 10 individuals containing number of wrongly and correctly predicted directions.*

Individual No.	No. of frames / Video	No. of Correctly Predicted Direction	No. of Wrongly predicted Direction	Total Accuracy %
1	61	57	4	93.44
2	29	29	0	100
3	31	28	3	90.32
4	30	29	1	96.66
5	25	24	1	96
6	23	22	1	95.65
7	89	87	2	97.75
8	51	46	5	90.19
9	68	68	0	100
10	35	30	5	85.71
Sum	**442**	**420**	**22**	**94.57**

As clear from the table, the average accuracy of this research is arrived up to 94.57%, where 22 attempts failed to achieve the correct recognition while 420 video frames are correctly predicted by the proposed control system. Although, the testing accuracy has been done on a recorded video, a real-time implementation can also be applicable for the proposed methodology. It is interesting to discuss the reason of resulted error, which is almost 5.5%, in order to be overcome by other researcher in the future, as in our perception, the problem might occur due to the colour skin detection method, which is in this research based on the YCbCr colour space because the illumination largely plays a role of the recognition performance. Another reason might undermines the recognition is that, if the originated hand shape is not so clear, that the algorithm get stuck trying recognizing, then resulting null. Figure 5 has three instances each one of them with its binary image to present how it is look after binary operation.

6. Conclusion

Four-control signals depending on vision-based hand gesture have been transferred correctively to a robot voice. However, this operation is very important to blind people, as well as has many applications in terms of remotely control on any electronic device. The proposed technique is achieved by detecting and segmenting the hand to give 4-directions as: up, down, left or right, as four directions either with free clockwise or anti-clockwise. The final feature vector extracted is the area of the hand object (convex). Then an expert system us used to predict the correct originated direction based on the area inference. In the future, any sign might modeled and used as a control signal to be functioned with any required control signal by using camera.

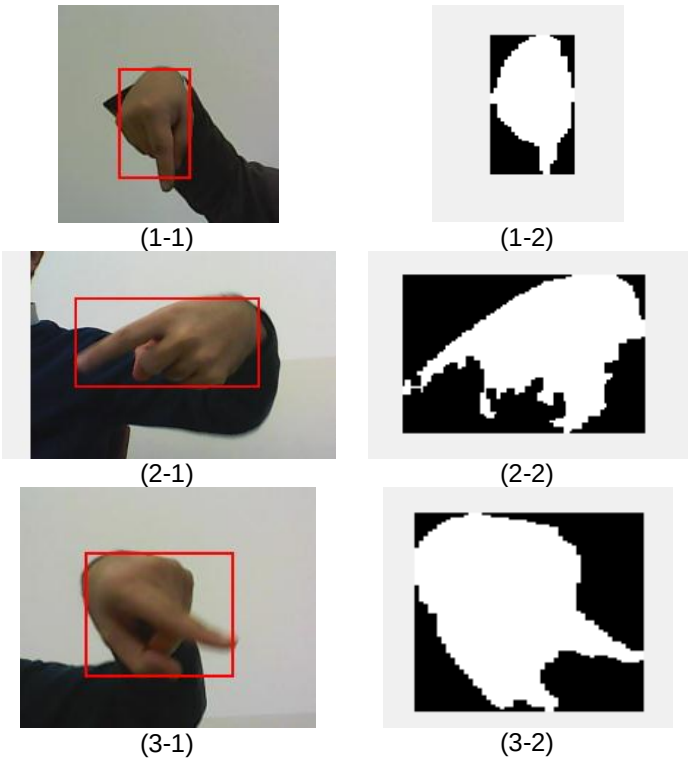


Fig. 5. *Illustrates random hand-gesture direction as incorrectly classified.*

REFERENCES

[1]. G. Murthy and R. Jadon, "A review of vision based hand gestures recognition," International Journal of Information Technology and Knowledge Management, vol. 2, pp. 405-410, 2009.

[2]. F. L. MALALLAH, et al., "HYBRID HAND-DIRECTIONAL GESTURES FOR BAIOMETRIC BASED ON AREA FEATURE EXTRACTION AND EXPERT SYSTEM," Journal of Theoretical & Applied Information Technology, vol. 95, 2017.

[3]. F. L. Malallah, et al., "A Review of Biometric Template Protection Techniques for Online Handwritten Signature Application," International Review on Computers and Software (I. RE. CO. S.), vol. 8, 2013.

[4]. F. L. Malallah, et al., "Online handwritten signature recognition by length normalization using up-sampling and down-sampling," International Journal

of Cyber-Security and Digital Forensics (IJCSDF), vol. 4, pp. 302-313, 2015.

[5]. S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," Artificial Intelligence Review, vol. 43, pp. 1-54, 2015.

[6]. N. Sae-Bae, et al., "Biometric-rich gestures: a novel approach to authentication on multi-touch devices," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2012, pp. 977-986.

[7]. N. Sae-Bae, et al., "Multitouch gesture-based authentication," IEEE transactions on information forensics and security, vol. 9, pp. 568-582, 2014.

[8]. M. S. M. Asaari and S. A. Suandi, "Hand gesture tracking system using Adaptive Kalman Filter," in 2010 10th International Conference on Intelligent Systems Design and Applications, 2010, pp. 166-171.

[9]. J. Doublet, et al., "Contactless hand recognition based on distribution estimation," in Biometrics Symposium, 2007, 2007, pp. 1-6.

[10]. A. Morales, et al., "Comparing infrared and visible illumination for contactless hand based biometric scheme," in 2008 42nd Annual IEEE International Carnahan Conference on Security Technology, 2008, pp. 191-197.

[11]. S. Berman and H. Stern, "Sensors for gesture recognition systems," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, pp. 277-290, 2012.

[12]. M. Chen, et al., "Air-Writing Recognition—Part I: Modeling and Recognition of Characters, Words, and Connecting Motions," IEEE Transactions on Human-Machine Systems, vol. 46, pp. 403-413, 2016.

[13]. K. Natarajan, et al., "Hand Gesture Controlled Drones: An Open Source Library," in Data Intelligence and Security (ICDIS), 2018 1st International Conference on, 2018, pp. 168-175.

[14]. Z. Zhao, et al., "Web-based interactive drone control using hand gesture," Review of Scientific Instruments, vol. 89, p. 014707, 2018.

[15]. Y. Y. Pang, et al., "A real time vision-based hand gesture interaction," in Mathematical/Analytical Modelling and Computer Simulation (AMS), 2010 Fourth Asia International Conference on, 2010, pp. 237-242.

[16]. F. Chen, et al., "Finger Angle-Based Hand Gesture Recognition for Smart Infrastructure Using Wearable Wrist-Worn Camera," Applied Sciences, vol. 8, p. 369, 2018.

[17]. Z. Qiu-yu, et al., "Hand Gesture Segmentation Method Based on YCbCr Color Space and K-Means Clustering," Interaction, vol. 8, pp. 106-116, 2015.

[18]. P. Jackson, "Introduction to expert systems," 1986.