

Human Distance Estimation using Quadcopter for Surveillance Purpose

Ahmad Zaki Shukor, Khaw Huai Jian, Fariz Ali@Ibrahim, Mohd Shahrieel Mohd Aras,
Muhammad Herman Jamaluddin

Abstract— Nowadays, quadcopters are commonly used. Quadcopters are unmanned aerial vehicles with four propellers to provide lift to fly and hover above ground. Quadcopter nowadays is a very common commercial item in everyday life. Some quadcopters are designed to do 3D or 2D mapping of a certain area or to take videos or just for entertainment purposes. Quadcopter is a very versatile item and is able to change into anything for example a quadcopter can also be used for security purposes to decrease the crime rate of our country. The objective of this study is to design and develop a quadcopter with image processing system to have the ability to measure the distance of a human from the drone itself. The quadcopter is designed to be small in size and have a mini computer like Raspberry Pi on top of it to compute the algorithm to calculate the distance of the human by using image processing technique through the camera which is setup on the drone. Human detecting algorithm YOLO and software Open CV is chosen to detect human and calculate the distance from the quadcopter. The results show that the system is quite limited by the capabilities of the hardware. The system shows an accuracy of more than 90 percent when the human is standing within a certain range. Both the accuracy of the distance sensing and human recognizing system is affected by the limitation of the hardware.

Keywords: Human distance estimation, quadcopter, surveillance, YOLO algorithm.

I. INTRODUCTION

Drone is one of the technologies with the most potential industry growing these days. Unmanned Aerial vehicles also known as drones are aircrafts that are navigated without any human pilot on board the vehicle [1]. Drones can either be navigated via control from ground or by using a Global Positioning System (GPS) tracking system. Drones come in different shape and sizes, some large drones that can carry up to 10kg of weight, some comes is a very small form factor such as DJI (Da Jiang Innovation) Spark. This drone is classified as a VTOL (Vertical Take-Off and Landing) drones which can take off, fly, hover and land vertically with the assist of gyro sensors and GPS system [2]. Companies like Amazon and Google also invested a lot in

this industry as drone industry is going to be one of the main industry in the future with its versatility and potential.

Residential break-ins are common and single family homes are the most frequent target. This is because these kinds of crimes are easy and generally non-confrontational and most of the deeds are done while the tenants are away from house. While it is not common to have any confrontation with the tenant, some burglars may encounter situations like these. When it happens the burglars will detain the residents and threaten them with violence. In cases that the tenants fight back usually it does not end well as the burglars are more prepared than the tenants are. Gated high-rise apartment complexes which are “equipped” with 24-hour guards and electronic access have a considerable lower crime rate than other housing but at the expense of a higher price point consider to other housing area [3]. There are reasons on why it is viable to choose a drone for surveillance purpose instead of the more traditional static surveillance camera. This is because drone can be used to follow the culprit instead of a static camera which can only be used at one place. Other than that, the operator can also fly the drone to desired places and have a better view of the area whereas there are a lot of blind spot of a static camera. If the culprit has done his homework and found the blind spot, then he can slip through with ease.

Use of sensors to detect intruders, such as ultrasonic range finders, infrared cameras, stereo vision camera or heat sensors might be able to successfully detect an object, but could not differentiate between a human subject and an object. Thus, it is even more challenging to differentiate the identity of a human detected. This is why image processing is one of the most preferred approaches nowadays, as many surveillance or security systems can identify a human subject based on certain features with high accuracy. However, real-time image processing is obviously affected by lighting and the view angle of the vision sensor.

Having a drone for surveillance may not be as easy as it seems. Aside from problems of power supply which affect flight time, the payload that a drone can carry also varies. Most static surveillance systems deploy state-of-the-art artificial intelligence image processing to identify a non-desired object/subject. For a mobile operating drone, it is even more difficult. This is because the drone has to be equipped with a high-power processing unit to perform real-time image processing, which incurs increase of payload on the drone. This is one of the reasons that there are not many DIY drones with artificial intelligence or image processing.

Revised Manuscript Received on September 14, 2019.

Ahmad Zaki Shukor, Mechatronics Department, Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia. (Email: zaki@utem.edu.my)

Khaw Huai Jian, Mechatronics Department, Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia.

Fariz Ali, Mechatronics Department, Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia.

Mohd Shahrieel Mohd Aras, Mechatronics Department, Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia.

Muhammad Herman Jamaluddin, Mechatronics Department, Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia..

Mostly are produced from companies with cutting-edge technology.

Having mentioned the difficulties of object and human differentiation, the next challenge would be to determine or estimate the distance between a human and the drone. This is necessary because the drone has to react to an incoming intruder either by keeping its distance of view and avoid from being attacked by the intruder, at the same time following the intruder. Stereo vision has proved feasibility of mobile robot navigation, car distance estimation for self-driving cars and could be a viable option for distance/depth estimation of human [12], [15], [16]. However, some other researchers shown that detection of human is achievable [13], [14] while others have shown that distance measurement could be successful by using radar sensors [17], [18].

In this research, we develop our own quadcopter equipped with a processing unit and assembled batteries for the experiments. Then, we tested its ability of flight by recording the voltage drop after flight time tests. Later we performed a human detection algorithm to identify a human and later estimated the distance of the human from the camera installed on the quadcopter. To verify the distance detection algorithm, we carried out two types of tests, static drone test and flight drone test.

II METHODOLOGY

There are two phases of the development of the drone in this research. The first phase is the hardware construction and testing and the second phase is the algorithm development and verification.

A Hardware Development

For the hardware development, the initial process was to choose the frame design. There are many types of frames used for quadcopters such as X frame, Plus frame and H frame [27]. Of the three, we chose X frame which has higher stability and easier to control. This is because the perpendicular distance between each motor is equal, thus having the same level of stability on all axis. For the material for the body of the quadcopter, some options available are fiberboard, aluminum, metal, or 3D printing material such as ABS or PLA. We chose PLA as it is much lighter and can be fabricated easily using a 3D printer.

The other components include four brushless motors with 960 kV rating, paired with 10 inch propellers with 3.8 inch pitch, with an upward thrust of 0.7kg and 2.8kg, an Electronic Speed Controller 20 Amps, a Raspberry Pi 3 B+ minicomputer to stream video via Wi-Fi, a 5MP Raspberry Pi Camera (set at 30 frames per second) mounted on a 3D printed bracket and batteries, as shown in Fig. 1.



(a)



(b)

Fig. 1.(a)The 3D drawing of the parts to be assembled and (b) the assembled quadcopter.

A flight payload test for the assembled drone was performed. The purpose was to assess the voltage drop incurred for a certain flight time. This is needed to ensure that the system has enough battery power to last for a few minutes. Three trials were done for different flight durations and the voltage drop, which is the difference between the voltages before and voltage after flight was measured by using a multimeter and recorded.

B Image Processing

Some quadcopter models include image processing units on the quadcopter itself. However, it is known that the processing power drains batteries quickly and might not be viable to last long. For us, we use the Raspberry Pi minicomputer to merely stream data via Wi-Fi in the form of video to a web address. This way, it saves the batteries of the drone, and images/videos are accessible via the internet.

An algorithm will be used to calculate and detect objects in the streamed video. This is done via the Personal Computer which has a higher processing power. The algorithm is YOLO (You Only Look Once) which 'looks' at the whole image once and then goes through the library or network once and detects/classifies the object. Other algorithms/frameworks usually look at different sections of the image multiple times at different scales and perform image classifications, which is relatively slow. YOLO processes the image faster. In our research, YOLO will be implemented by using the OpenCV library. OpenCV provides a common infrastructure for computer vision, and provides the graphical user interface that can run YOLO. Programming was done in Python language.

YOLO algorithm uses a single convolutional network which predicts multiple bounding boxes and class probabilities at the same time. It uses 24 convolutional layers and 2 fully connected layers, alternating 1 x 1 convolutional layers to reduce the feature space from preceding layers. It trains on full images and optimizes performance. Detection is treated as a regression problem and shows 45 frames per second speed on a Titan X GPU. Streaming video can be processed in real-time with less than 25 milliseconds of latency. The YOLO algorithm divides the input image into S x S grid. In the condition that the center of an object lies in a grid cell, that grid cell is responsible to detect that object. Each grid cell predicts bounding boxes, B and confidence scores for the boxes. The confidence scores show how confident the box contains an object and how accurate is the box it predicts.

Confidence is defined as $Pr(Class_i | Object) * Pr(Object)$. If there is no object in that cell, the score of confidence should be zero. Otherwise, the confidence score is equal to the intersection over union (IOU) between predicted box and the ground truth. Each box contains 5 predictions; x,y,w,h and confidence/ The (x,y) coordinates represent the center of the box relative to the grid cell bounds. The width and height are predicted relative to the whole image.

Each grid cell also predicts C conditional class probabilities, $Pr(Class_i | Object)$. These probabilities are conditioned on the grid cell containing an object. Only one set of class probabilities per grid cell is predicted, regardless of the number of boxes B. At test time, the conditional class probabilities and the individual box confidence predictions are multiplied as in (1)

$$Pr(Class_i | Object) * Pr(Object) * IOU_{truth}^{predict} = Pr(Class_i) * IOU_{predict}^{predict} \quad (1)$$

which shows class-specific confidence scores for each box. These scores encode both the probability of that class appearing in the box and how well the predicted box fits the object.

Before implementing the algorithm in real-time, calibration of the camera is needed. A setup of the drone on a static surface is needed to ensure the y-pixel values are estimated, as the algorithm will construct a bounding box on the identified human subject. The human subject will stand in front of the drone, for a distance of 3m to 12m, as in Fig. 2. The acquired y-pixel values are recorded for each increase of 1m of distance. The calibration is repeated twice to obtain the average reading.



Fig. 2. Calibration of the drone to obtain a best-fit line for distance estimation.

From the data of pixels of different distances, a graph is plotted to obtain the best fit line. This will result in an equation to be used later by the real-time algorithm

C Human Distance Estimation

Next, the human distance estimation was experimented for stationary drone and another test when the drone is in actual flight. The calibration result as depicted in Fig. 3 obtains the best-fit line from the average of two calibration cycles, was used to estimate the distance, i.e. a certain pixel value of the bounding box will indicate the calculated/estimated distance.

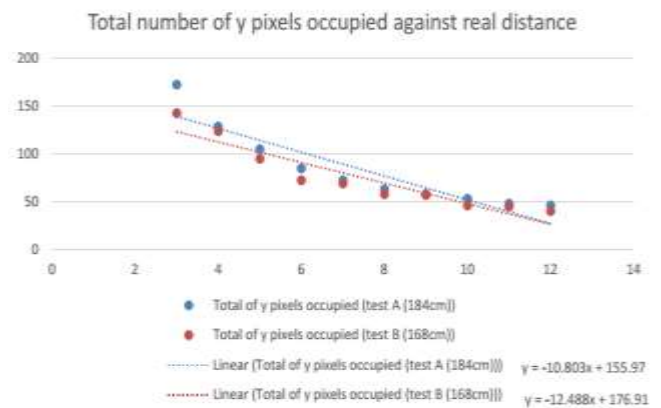


Fig. 3. Best fit-line for distance estimation.

The first estimation was done in the stationary mode, where the system is held stationary at a certain height by a human. The setup for the stationary test is illustrated in Fig. 4. The distance algorithm is run in real-time and screenshots were taken of the processed image and estimated distance. A human will hold the drone and stay static while the tested human subject will move from a distance of between 3m to 12m, with an increase of 1m (measured by a tape placed on the ground). Three trials were performed, and the average distance estimated is calculated to obtain the error for the distance estimation.



Fig. 4. Stationary mode test.

The other experiment was intended to assess the accuracy of the distance estimation in actual flight. The setup is shown in Fig. 5. The pilot flies the quadcopter, while the tester (human subject) stands in front of the drone, walking the distance from 3m to 12m, with increments of 1m each time. Three trial experiments were carried out and the average distance was calculated to assess the accuracy or percentage of error.



Fig. 5. Flight mode test.

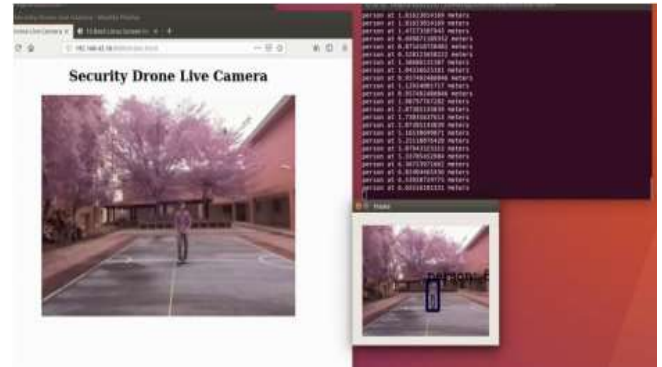


Fig. 6. Outdoor test.

In addition, several different poses of the human was tested for the system; facing the camera and standing, ducking, different human yaw angles with respect to the camera at 0,30, 60, 120 and 180 degrees, lying flat on the ground, walking motion and running motion. These tests are important because most of the time, the human subject is in motion and at different positions/poses. The system should be robust enough the detect the human presence at whatever pose.

III RESULTS AND DISCUSSION

The first test of payload was intended to investigate the voltage drop of the quadcopter after 10 minutes, 20 minutes and 30 minutes of flight time. This is to ensure that the quadcopter will be able to last a certain amount of time in the air. Each flight time was tested three times. The weight of the whole system (quadcopter, batteries and other components) were 175 grams in total. The results are depicted in Table I

Table- I: Voltage drop for flight times of 10, 20, 30 minutes

Flight Time	Voltage Drop		
	Trial 1	Trial 2	Trial 3
10 minutes	0.93	0.89	0.95
20 minutes	1.89	1.82	1.86
30 minutes	2.77	2.87	2.85

The results in Table I show that the voltage dropped for every 10 minutes throughout all the trial runs has a consistent drop of around 0.9 volt per 10 minutes. For a 3S battery (12.6 volt) it is recommended to be used until the voltage drops to 9 volt which is a 3 volt drop to ensure the battery can be recharged and still in good condition after use. The table shows that even with the weight of the system on the drone it can still manage a flight time of 30 minutes. This is an acceptable flight time considering that most of the commercial drone nowadays can fly for 15 to 20 minutes.

For the stationary test, the ground station screen is shown in Fig. 6. The picture on the left shows the web browser displaying the video stream of the camera, the upper right picture shows the display of distance estimated and lower right window shows the identified human in the bounding box. The results of the experiment of distance estimation for a stationary drone are shown in Fig. 7.

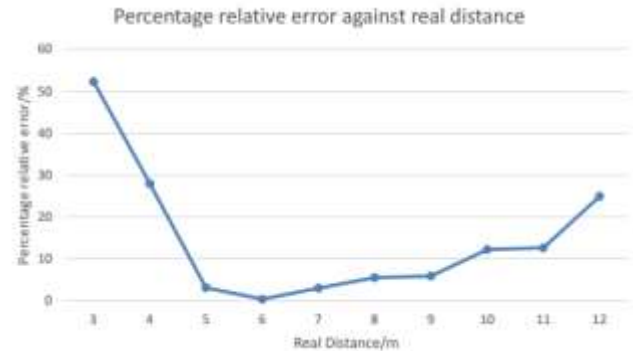


Fig. 7. Results of the outdoor stationary test.

The results from Fig. 7 show that the error starting at 3 meters is very high then it becomes lower and lower as the tester is further and further away from the system until the tester is 7 meters then the error will start to increase again. This is because as the tester moves further and further away from the system the changes in y axis pixel is decreasing until it is so small that the system cannot differentiate it. The accuracy from 5 to 7 meters is very high and is the desirable results. One of the factors affecting the accuracy is because of the processing power of the Raspberry Pi and the ground station. For this setup, the Raspberry Pi can only stream video of 320x240 pixels to the ground station and the ground station is only capable of processing the video at 8 frames per seconds. If the video resolution is increased the ground station will not be able to process it at a reasonable speed. If the processing power of the ground station is increased the system can have a higher resolution and thus a more accurate data can be obtained. The result of the distance estimation during flight is shown in Fig. 8.

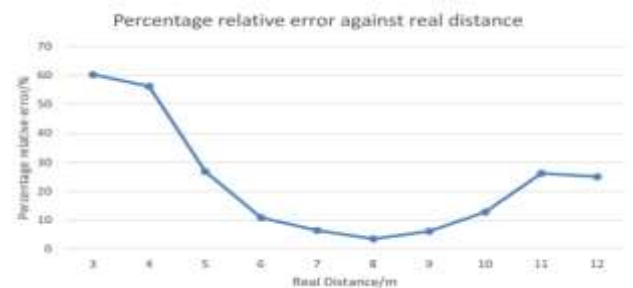


Fig. 8. Results of the outdoor flight mode test.

From Fig. 8 we can see a similar trend to the stationary drone experiment. The error will decrease until a certain point and then increase again when the tester is too far away. This is due to the same reasons mentioned in the previous discussion. But for this experiment it is noted that the relative error is higher than the previous experiment. This is because this drone does not have an onboard GPS. The onboard GPS can allow it to fly in loiter mode. This mode allows the drones to hover in place without the control of the pilot. Because this drone does not have an onboard GPS, it causes the drone to hover and move around and the pilot has to constantly fly it back to the original place. The increase in error is due to the fact that the drone is moving.

For the different pose tests, the snapshots are shown in Fig. 9. The system succeeded to identify for all conditions (i.e. facing camera, ducking, and yaw angle) except for the pose where the human subject lies on the ground or the subject runs past the drone. It is logically difficult to identify because of the flat position of the human subject and also difficult when the speed of the subject motion is very fast.

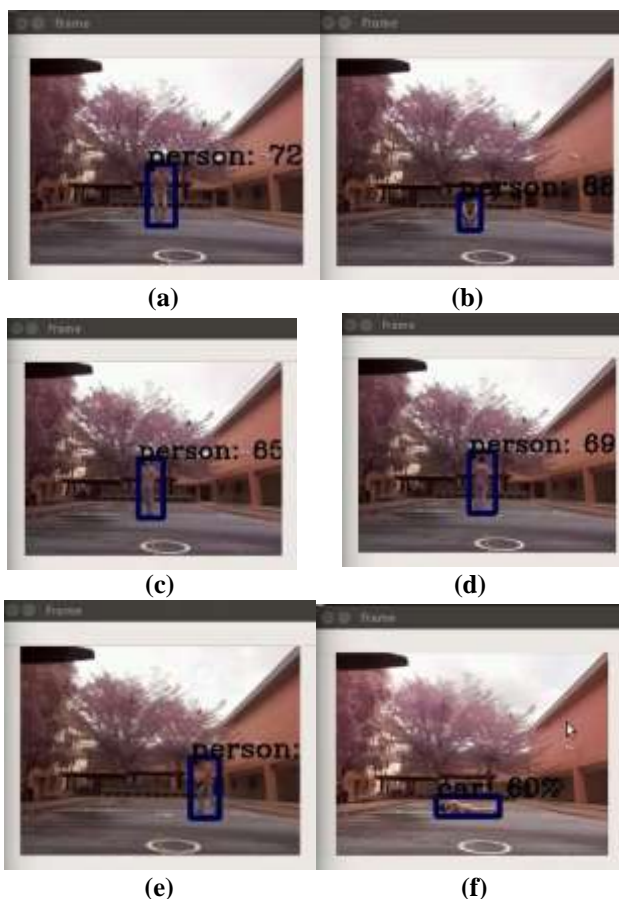


Fig. 9. Human identification results at different poses; (a) standing and facing camera, (b) ducking, (c) yaw angle 30; (d) yaw angle 180, (e) walking, and (f) lying down.

IV CONCLUSION

The human distance estimation was implemented in the view of the camera that was placed on the quadcopter, using the YOLO algorithm which constructs a bounding box around the human identified. Initial calibration was needed to relate the human distance with the pixel values of the bounding box, which was done by using regression, obtaining the best fit line to relate the two variables. The

drone was experimented on two different setups; stationary drone and flying drone. The error of distance estimation in the flying drone setup was higher due to the slight motion of the drone. In both setups, it was found that for a distance of 3m to 5m, the error is quite high due to the large pixel values as the bounding box of the human becomes large. Errors were also higher from 10m to 12m as the pixel values becomes low due to the bounding box size that decreases.

Even when the human is turning around (incurring a yaw angle change), the system was still able to identify the human, except in the case of high motion (running) and lying down. This is due to the limitation of the camera capture speed and image processing time, since the human runs at high speed. When lying down, the system identifies the object as a car, instead of a human, which is false.

In future, it is desired to investigate the effects of distance estimation when the drone is moving around and not hovering at a certain location, which is deemed to be more difficult. A higher frame rate capture setting for the Raspberry Pi camera (or a faster camera) is also needed to be able to identify the human in a high speed motion (i.e. running). This research could benefit in the way that drones can be used for surveillance and security purposes, which could assist human in the said industry. Looking ahead, drones could also be used to follow or track a moving person in a vicinity.

V ACKNOWLEDGMENT

We would like to acknowledge the research group, Robotics and Industrial Automation under the Centre for Robotics and Industrial Automation (CeRIA). We would also like to acknowledge Universiti Teknikal Malaysia Melaka (UTeM) in general for the facilities and resources provided.

REFERENCES

1. F. Corrigan. (2019). How do drones work and what is drone technology. [Online]. Available: <https://www.dronezon.com/learn-about-dronesquadcopters/what-is-drone-technology-or-how-does-drone-technology-work/>.
2. Hobbytron. (2018). What are drones. [Online]. Available: <https://www.hobbytron.com/lc/what-are-drones.html>.
3. D. Joshi. (2017). Commercial Unmanned Aerial Vehicle (UAV) market analysis – Industry trends, companies and what you should know. [Online]. Available: <https://www.businessinsider.com/commercial-uav-market-analysis-2017-8/?IR=T>.
4. D. Murray and J. J. Little, "Using real-time stereo vision for mobile robot navigation," *Auton. Robots*, 8(2), 2000, pp. 161–171.
5. Y. M. Mustafah, R. Noor, H. Hasbi, and A. W. Azma, "Stereo vision images processing for real-time object distance and size measurements," *Int. Conf. Comput. Commun. Eng.*, 2012, pp. 659–663.
6. Y. D. Salman, K. R. Ku-mahamud, and E. Kamioka, "Distance measurement for self-driving cars using stereo camera," *6th Int. Conf. Comput. Informations*, 2017, pp. 235–242.

7. Y. S. Kim, J. Il Park, D. J. Lee, and M. G. Chun, "Real time detection of moving human based on digital image processing," SICE Annu. Conf., 2007, pp. 2030–2033.
8. E. Corvee, S. Bak, and F. Bremond, "People detection and re-identification for multi surveillance cameras," Int. Conf. Comput. Vis. Theory Appl., 2012, pp. 82–88.
9. M. Z. Ikram, A. Ahmad, and D. Wang, "High-accuracy distance measurement using millimeter-wave radar," IEEE Radar Conf., 2018, pp. 1296–1300.
10. S. Ayhan, S. Scherr, P. Pahl, T. Kayser, M. Pauli, and T. Zwick, "High-accuracy range detection radar sensor for hydraulic cylinders," IEEE Sens. J., 14(3), 2014, pp. 734–746.



Muhammad Herman Jamaludin received the B.E. from Universiti Teknologi Malaysia in 2001 and M.Sc. from Universiti Teknikal Malaysia Melaka in 2009. He obtained a PhD degree in Yokohama National University, Japan in 2015. Since 2003, he has been attached with the Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, where he is currently an Associate Professor. His research interests include haptic, robotics, and embedded system. He is a graduate member of the Board of Engineers Malaysia.

AUTHORS PROFILE



Ahmad Zaki Shukor received the B.E. from Universiti Teknologi Malaysia in 2002 and M.E. from University of South Australia in 2004. He obtained a PhD degree in Yokohama National University, Japan in 2013. Since 2002, he has been attached with the Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, where he is currently an Associate Professor. His research interests include actuators, robotics, and motion control. He is a Professional Engineer of the Board of Engineers Malaysia



Khaw Huai Jian obtained his Bachelor of Mechatronics Engineering in Universiti Teknikal Malaysia Melaka in 2019. His research interests include quadcopters flight control, embedded systems and robotics.



Fariz Ali Obtained his highest education in Yokohama National University when he did his Doctor of Engineering in field of robotics in 2013 under supervision of Professor Atsuo Kawamura, the robotics expert from Japan. During his Phd study, Dr Fariz has develop the walking algorithm for bipedal walking robot by using Dual Length Linear Inverted Pendulum (DLLIPM) method. Earlier, Dr. Fariz has completed his master degree in University of South Australia, Adelaide in 2002-2004. He completed his undergraduate study at Universiti Teknologi Malaysia in mechatronics engineering from 1998 to 2001. He is currently working in Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka as senior lecturer. His research interest is humanoid robot, 6-legged robot, and other mechatronic project. He has successfully supervised 3 postgraduate students and about 25 undergraduate students for their final year projects. Dr Fariz has won three awards during the 26th International Invention, Innovation and Technology Exhibition 2015 and one of the product is known as "RH2000 Cybernatics Hand". Dr Fariz has published articles more than 19 journal papers and 12 conference papers. In 2007, the Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka has recognized Dr. Fariz's contributions and services by giving him an award known as "Anugerah Perkhidmatan Cemerlang" which is the excellence award for academician.



Mohd Shahrriel Mohd Aras is Associate Professor at Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka UTeM. His current research is focusing on control system design of underwater technology. His primary interests related to underwater robotics and Artificial Intelligence.