❏   528

Multiple Human Body Postures Detection using Kinect

Rosdiyana Samad[*], Law Wen Yan, Mahfuzah Mustafa, Nor Rul Hasma Abdullah, Dwi Pebrianti
Faculty of Electrical and Electronics Engineering, University Malaysia Pahang, Malaysia

Article Info	ABSTRACT
	This paper presents a method to detect multiple human body postures using Kinect sensor. In this study, a combination of shape features and body joint points are used as input features. The Kinect sensor which used infrared camera to produce a depth image is suitable to be used in an environment that has varying lighting conditions. The method for human detection is done by processing the depth image and joint data (skeleton) which able to overcome several problems such as cluttered background, various articulated poses, and change in color and illumination. Then, the body joint coordinates found on the object are used to calculate the body proportion ratio. In the experiment, the average body proportions from three body parts are obtained to verify the suitableness of golden ratio usage in this work. Finally, the measured body proportion is compared with Golden Ratio to determine whether the found object is a real human body or not. This method is tested for various scenarios, where true positive human detection is high for various postures. This method able to detect a human body in low lighting and dark room. The average body proportions obtained from the experiment show that the value is close to the golden ratio value.

Corresponding Author:

Rosdiyana Samad,
Faculty of Electrical & Electronics Engineering,
Universiti Malaysia Pahang,
Pekan, 26600 Pahang, Malaysia.
Email: rosdiyana@ump.edu.my

1. INTRODUCTION

During the last decade, human detection has gained huge attention in computer vision and pattern recognition research communities due to the variety of applications that can be made by implementing human detection in their system [1]. Human detection is an essential component and challenging in surveillance application, abnormal event detection, fall detection, human gait characterization, healthcare application and etc. Nowadays, many applications have been made to operate in real-time. However, to detect the human in real-time is not an easy task. A lot of issues should be considered during the human detection development. One of the challenges is the non-rigid nature of the human body that creates numerous possible poses. The human body pose is always changing whenever human moves. Normally, the detection works successfully when a person is standing in an upright position, but the detection may lose the target body when that person moves to another position, such as sit down on a chair or bending the body. Human is always moving in the video and the size of the body varies when the position and direction of the camera is changed. Sometime, there is a cluttered background often occurs in the outdoor scene. Moreover, occlusion that can caused by multiple humans moving in the crowded scene or other things close to human also occurred. For the indoor environment, the problem is arisen due to illumination change. Different room locations would give different lighting, based on the lamp and condition at that time. For the outdoor environment, the weather changes also caused the varying illumination and it make the human object become hard to detect.

In recent literature studies shown that many researchers are interested in human detection studies and various methods have been developed as shown in [2]. Extracting the human body is an essential and crucial process in the human detection method. Many challenging issues can be arisen in the human features extraction method. Therefore, an appropriate technique should be selected carefully so that the whole detection process will produce a better accuracy and the detection is reliable/robust. In feature extraction method, there are a number of ways to extract the human body region and a few types of features will be produced, such as motion, appearance, shape and combination features [1]. In previous studies showed that shape feature is much preferred by the researchers compared to other type of features. Shape features can be described as object location, orientation, the edge information, pixel intensity or binary contours represent the human shape. To represent the human descriptor, the shape features also can be combined with various features. Different types of features can generate different information to the descriptors and the descriptions of the human object are more distinguishable in various viewpoints and poses.

In object detection study, Y. Mustafah et al. [3] used stereo image for real-time object distance and size measurement. Their results showed that stereo image has a considerable accuracy for object detection. However, one of the drawbacks in their research is the used RGB image for object detection required constant environment lighting. Secondly, the proposed method used two cameras which mean they need an extra cost for extra cameras. Thirdly, the accuracy for object detection depended on the image resolution. Therefore, high processing power computer is most likely required in order to process the high image resolution image in real-time. Other researchers, L. Zhang and Y. Liang [4] used background subtraction method to track moving object. The authors minimized the drawbacks of using RGB image such as the need of constant environment lighting in [3] by updating the image background in real time. Shortcomings in the traditional method of object detection such as adding or removing objects from the background are also dealt with to an extent by applying threshold. The setbacks will most likely be due to the use of RGB image such as sudden changes in the environment lighting (e.g. from bright to pitch black) and detection in very bright or pitch black room.

The use of sensor in object or human detection application has to be chosen carefully. Different types of camera normally give a different result. M. Smisek et al. [5] experimentally investigated the measurement of the depth resolution and error properties of Kinect sensor. The author also made a quantitative comparison of the 3-D measurement capability for Kinect sensor, two medium Nikon D60 SLR cameras in stereo rig and SwissRanger SR-4000 3D-TOF camera. The results show that SLR Stereo was the most accurate, close by Kinect sensor and SR-4000 3D-TOF being the least accurate. Other researchers, T. Stoyanov et al. [6] also compared Kinect sensor, SwissRanger SR-4000 camera and Fotonic B70 TOF camera to a standard actuated laser range finders (aLRF). The evaluation was carried out with known ground truth data produced by aLRF and in an uncontrolled environment. The results show that the performance of Kinect sensor was very close to the laser sensor for short range environments with distance less than 3.5 meter, the two TOF cameras had slightly worse performance in the short range test and no sensor achieved performance comparable to the laser sensor at full distance range. From these researches, it found that the Kinect sensor is suitable to be used in human motion detection application, especially for the indoor environment. Other researchers in [7-11] also used Kinect sensor as an imaging device for the gesture recognition, robot control system and rehabilitation.

In this paper, a method for detecting multiple human body postures and poses using Kinect is proposed. This human detection algorithm is developed by using C++ programming, OpenCV and NiTE libraries. The proposed method uses a combination of shape features that represented by contours and skeleton (that represented by joint points). In this method, the Golden Ratio, ϕ is used in the body part ratio measurement to discern whether the shape of the found object is human or not. In this study, the algorithm has been tested to detect a person that moving around in an indoor environment, with variations of postures and poses. The Golden Ratio is the ratio of the sum of the quantities of the larger quantity is equal to the ratio of the larger quantity to the smaller one [12]. It has been used in many applications such as plastic surgery simulation software, animation software, art, architecture and anatomy [12]. The Golden Ratio also has already been found in many physical, natural and human fractal structures [13]. In another study [14] shows that the value of the human gait ratio is close to Golden Ratio, $\phi \approx 1.618034$ in healthy subject.

2. RESEARCH METHODOLOGY

To develop human detection, the process is divided into three stages, which are body's contour detection, body joint detection and decision making. Nine subjects have participated in the experiment to get the body part proportions. The measured body part proportion is used in the decision making stage to verify the suitableness of golden ratio usage

2.1. Body Contour Detection

Connection with Kinect sensor is established in order to obtain RGB image, depth image and joint data from the sensor. After OpenNI and NITE are initialized, several parameters such as resolution, frame per second and pixel format are set before any data is obtained from the Kinect sensor. Both RGB and depth image are obtained from the Kinect sensor after the initialization. These images are then converted from RGB to BGR as OpenCV process color image in BGR format. Figure 1 shows the converted RGB and depth images. The depth image is displayed in gray scale where a brighter pixel indicated the further away from the Kinect sensor.

After the conversion, the background segmentation with improved adaptive Gaussian mixture algorithm [15] is applied to the depth image. Background subtraction is an approach to detect moving objects from the difference between the current frame and the background model. However, background subtraction based on the static background model is not applicable in real environments when other factors such as moving objects, shadows and various lighting condition leads to background changes. In order to adapt to the changes, each background pixel is modelled with a mixture of appropriate number of Gaussian distribution. The time proportions that those pixels stay in the scene is represented the weights of the mixture. The pixels which stay longer and more static are probably the background pixels. This method is an improvement from the first method that introduced by Z. Zikovic et. al [16] and it provides better adaptability to varying scene due illumination changes.

In this algorithm, four parameters have been set, which are the length of motion history of the foreground object, shadow detection, threshold value and learning rate. The learning rate indicates how fast the background model is learned and its formula is shown in Equation 1. The learning rate with value between 0 and 1 indicates how fast the background model is learned where 0 means the background model is not updated at all and 1 means the background model is completely reinitialized from the last frame. If the value of the learning rate is negative, the learning rate will use the value of history with calculation as shown in Equation 1.

The length of motion history is set as default value, the threshold is set to 16, the shadow detection is disable and learning rate is set to -1. The learning rate is set low so that the object of interest (e.g. people) is not absorbed into the background model when the object enters the scene and stops for a few seconds, whereas allow changes such as adding or removing furniture to be absorbed into the background model after some time. In this study, there is no difference whether the shadow detection is set to true or false because the depth image used in this work only require shape information but not a shadow.

$$learning\ rate = \frac{1}{\min(2 \times n\ frames, history)} \qquad (1)$$

Figure 2 shows the output of background subtraction applied to depth image when an object (human as the object as shown in Figure 2) enters the scene. White pixels are assigned to the difference between the current frame and the background model. Excluding the object, the white pixels around the image are probably caused by noise. Background subtraction is used in this system for two reasons. Firstly, it allows the humanlike object to be excluded in the detection such as the mannequin. No ROI will be set for the mannequin if the mannequin is already in place before the system is initialized. Secondly, the background subtraction algorithm in OpenCV allows changes in the surrounding to be updated to the background model. In case a mannequin is placed into the scene after the system is initialized, the mannequin will slowly be absorbed into the background model as time passes.
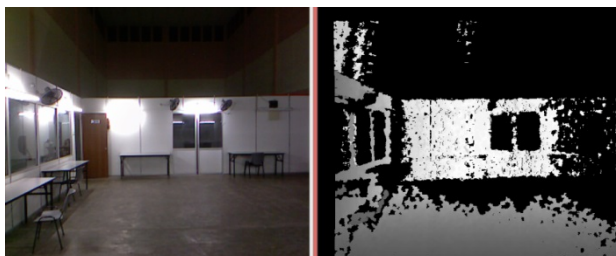

Figure 1(a) RGB Image after Converted from RGB to BGR format and (b) Remap Depth Image Displayed in Grayscale.

Figure 2. Output of Background Subtraction Applied to Depth Image when an Object Enters the Scene.

After background subtraction is done, distance transform is applied to calculate the distance to the closest zero pixel for each pixel in the source image. The used algorithm is described in [17] and [18]. The distance image is normalized with alpha and beta are set as 0 and 1. In this work, normalization is then applied to visualize the image with the lower and upper range boundaries, alpha and beta, are set at 0 and 1. After that threshold is applied to the normalize image. Figure 3(a) and Figure 3(b) show the output after applied distance transforms and the normalized image. The normalized image gives a better visual representation compare to the distance transform result.

Next, the normalized distance transform image is converted to a binary image with a threshold value of 0.5 as shown in Figure 4. This process is used to retain the shape of the body while removing hands, legs, and other objects that are not related to the body. The contour of the body in binary image is detected and drawn on RGB image as shown in Figure 5. Some unwanted objects which remain from the threshold process have to be removed. These contours are eliminated by comparing its contours with the predefined contour size which is 150 pixels. If its contour is smaller than the predefined contour size, the pixel (1- white) is replaced with 0 value that represent black color. To make the body detection appearance more appropriate, the region of interest is set as a red bounding box around the contour in Figure 6. The region of interest is used in a latter section to detect human presence.
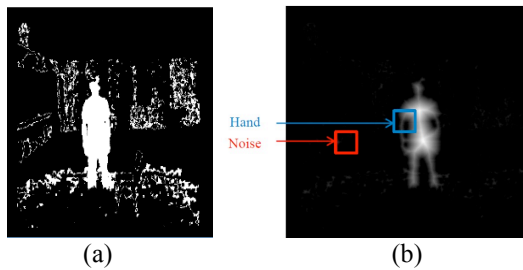

(a)	(b)

Figure 3 (a) Distance Transforms Output and (b)normalized image of a human body

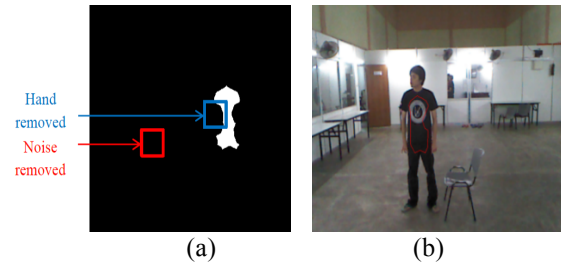

(a)	(b)

Figure 4. Threshold image Figure 5. Contour of the body is detected

2.2. Body Joints Detection

The skeleton tracking algorithm from NiTE 2 library [19] is used in this work to detect the joints. In the skeleton tracking, the body parts are represented by 15 points of the joint [20]. Joint data are obtained after the Kinect's initialization that has been explained previously. The joint data represent the location and distance of the subject from the Kinect sensor in millimeter. In this part, conversion is needed to draw the detected points and skeleton on the RGB image as shown in Figure 7.

2.3. Decision Making

A comparison has been made to decide whether the contour of the body is correct belongs to human's body. If the joint point is located within the region of interest (the rectangular box) as shown in Figure 7, the selected length between joints is calculated using a distance formula as shown in Equation 2.

$$length = \sqrt{(x1 - x2)^2 + (y1 - y2)^2 + (z1 - z2)^2} \qquad (2)$$

The proportion, P in human body as shown in Figure 7 is calculated using Equation 3, where r is a length of the right hand to right elbow, while t is length from neck to torso. The proportion is represented by coefficient P in the proportional fit equation as shown in Equation 4. In this work, four lengths are calculated to find the ratio or proportion of each part and these selected lengths are listed in Table 1. The ratio value is used to make a comparison with the Golden Ratio φ value. The length from neck to torso, t is used as base as this length can be obtained more accurately when a person is facing at another angle.

$$Proportion, P = \frac{r+t}{t} \qquad (3)$$

$$y = P \times x \qquad (4)$$

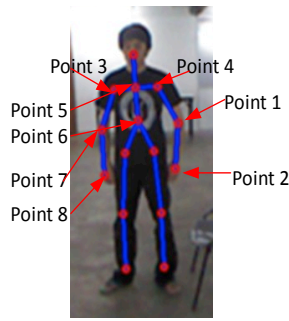Figure 6. Region of Interest in Red Rectangle based on the Contour Found

Figure 7. Selected Lengths between 2 Joints

Table. 1 Selected length for proportion calculation

Point no.	Label	Length measurement
Point 1 →Point 2	r	Right hand to right elbow
Point 3→ Point 4	s	Left shoulder to right shoulder
Point 5→Point 6	t	Neck to torso
Point 7 →Point 8	l	Left hand to left elbow

To determine the object is a real human, golden ratio (φ≅1.618) can be an excellent tool in estimating the proportions of the human body, the model of the human body is created based on this to estimate the size and proportion of the human. In this work, the golden ratio value is used to calculate the percentage of the proportion as shown in Equation 3. If the percentage of the proportion is greater than 80%, the detection is positive, which mean it is a real human.

$$Proportion\ percentage\ (\%) = \left(1 - \left| \frac{1.618 - P}{1.618} \right| \right) \times 100 \qquad (4)$$

In this experiment, nine subjects performed various movements in the room at one time. The body detection and skeleton tracking are performed and the length measurements of each body part are recorded. After that the ratios of the lengths are calculated. The Kinect performance also tested, by observing the effective area for the detection.

3. RESULTS AND DISCUSSION

Table 2. shows the results of the proportion calculations for three body parts of nine subjects. The average ratios for each body part are shown in Table 2, which are 1.644, 1.685 and 1.678. When these results are compared with the golden ratio value (1.618), the results show that the average ratios of three body parts are nearly close to the golden ratio value. From the comparison, it shows that the golden ratio can be used to estimate the human body detection based on the human body proportion.

Table 2. Calculated Human Body part Ratios

Subjects	Ratio between length from left shoulder to right shoulder and length from neck to torso (s:t)	Ratio between length from left hand to left elbow and length from neck to torso (l : t)	Ratio between length from right hand to right elbow and length from neck to torso (r : t)
1	1.709	1.685	1.671
2	1.595	1.699	1.823
3	1.587	1.664	1.637
4	1.687	1.672	1.648
5	1.679	1.756	1.780
6	1.602	1.565	1.594
7	1.582	1.609	1.597
8	1.658	1.887	1.715
9	1.693	1.624	1.638
Average	1.644	1.685	1.678

Figure 8 and Figure 9 show the results for true positive of human detection in a few postures and room condition. The proposed detection system successful to detect a subject that standing upright, facing backward, sitting on the floor and sitting on a chair as shown in Figure 8 (a), (b), (c) and (d). Meanwhile, this system also can detect multiple subjects in the scene as shown in Figure 8 (e). The Kinect works well in the

bright room and dark room. This is because its infrared projector and sensor able to capture a depth image, although the lighting in the surrounding area is totally dark. Figure 9 shows the true positive human detection when the room light is off.

In this work, the body detection is not only used the information from the output of the distance transform, but it is combined with the percentage of body proportion, to validate that the detected object is a real body. This body proportion can be used to differentiate between human and other living things. The body proportion is based on joint data from the skeleton tracking, and it will not be affected by the lighting changes. Many articulated poses can be detected as the body proportion ratio will not change with posture and movement. If shape information can be obtained accurately, the detection also can be performed in different level of terrain such as detecting a person walking on stairs. This body proportion ratio may be used as a human feature to differentiate between human and other living things.

Figure 10 and Figure 11 show the results for false negative detection when a subject is at the edge of the horizontal field of view and located more than 4 meters away from the Kinect sensor. Even though the horizontal field of view is approximately 60°, the joint data can only be obtained when the whole body is located in 50° of the horizontal field of view. In Figure 10, it is clear that a part of the body is not detected in the depth image. This system unsuccessful to detect a subject that is lying on the floor. When there is a continuous depth value between the body and the object as shown in Figure 12, the detector may falsely recognize the body and the object as a single object.
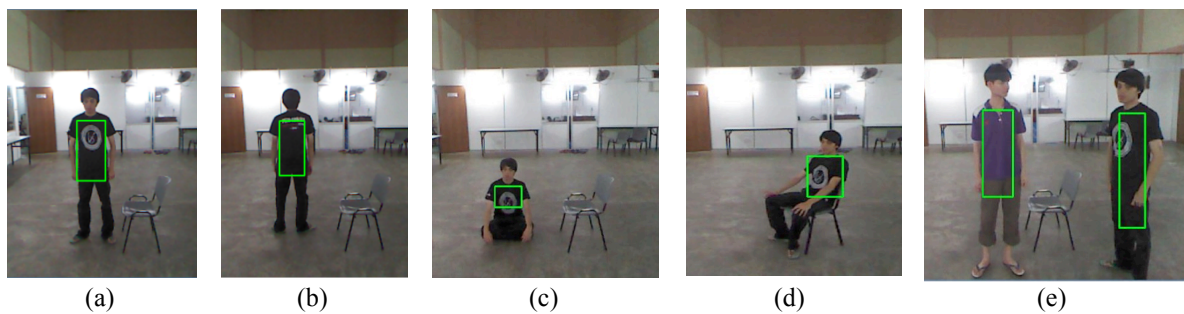

| (a) | (b) | (c) | (d) | (e) |

Figure 8. True Positive Detection when a Subject is (a) Standing Upright, (b) Facing Backward, (c) Sitting on the Floor, (d) Sitting on a Chair and (e) Multiple Subjects in the Scene.

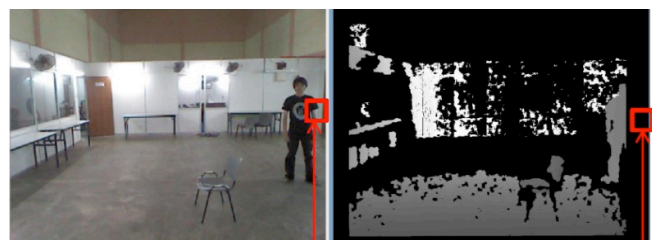Figure 9. True Positive Detection When the Room Light is off.

Figure 10. False Negative Detection when a Person is at the Edge of the Horizontal Field of View.
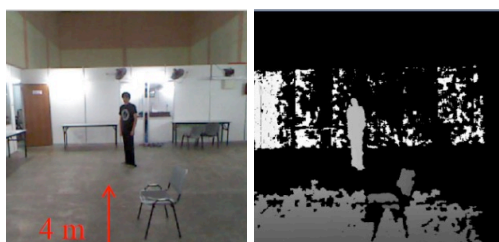

Figure 11. False Negative Detection when a Person is More than 4 Meters Away from the Kinect Sensor.
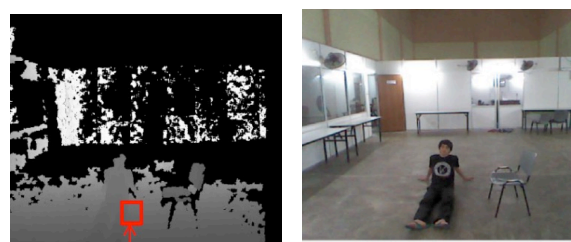
Figure 12. False Negative Detection when a Person is Lying on the Floor.

Figure 13 shows the Kinect effective range is from 1.5 to 4 meters with 20° horizontal field of view. The Kinect is not so effective from 1 to 1.5 meters and, 4 to 4.5 meters and between 50° to 60°. For the rest, the Kinect is not able to detect at all. Since the detection is based on the upper part of the body, the detection may work for close range as long as joint data can be obtained from the upper part of the body. When a subject is more than 4 meters away from the Kinect sensor as shown in Figure 13, the joint data cannot be obtained. This system failed to measure the body proportion when the subject is perpendicular to the Kinect sensor and when a subject is holding an object.
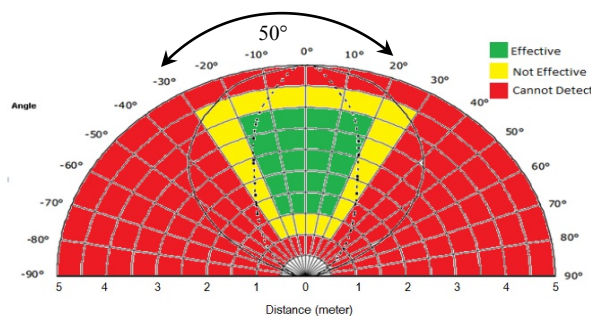

Figure 13. Kinect Effective Range

4.   CONCLUSION

In conclusion, the human detection system was successfully developed using Kinect sensor. The used of shape features together with the joint data improves this detection system by comparing the body proportion ratio of the found object with the golden ratio value. This system is able to run in real-time and having better performance in various illumination environments. The detection is successful to perform in low lighting and darker room. This detection method is able to detect human body for various articulated poses and multiple people in the scene. However, this system is unable to detect when a person was closely attached to an object such as lying on the floor and leaning against a wall. In this study also discover that the measured body proportion ratios from three parts of the body are almost near to golden ratio value which is 1.618.

5.   ACKNOWLEDGEMENTS

REFERENCES

[1]   D. T. Nguyen, *et al*., "Human detection from images and videos: A survey," *Pattern Recognition*, vol. 51, 2016, pp. 148-175.
[2]   B. Li, *et al*., "Indoor Human Detection using RGB-D images," in *Proc. of the IEEE Int. Conference on Information and Automation*, Ningbo, China, 2016, pp.1354-1360.
[3]   Y. Mustafah, *et al*., "Stereo Vision Images Processing for Real-time," in *Proc. of the Int. Conf. ICCCE*, 2012, pp. 659-663.
[4]   L. Zhang, and Y. Liang, "Motion human detection based on background," in *Proc. of the 2nd Int. Workshop ETCS*, vol. 1, 2010, pp. 285-287.
[5]   J. Smisek, *et al.,* "3D with Kinect," in *Proc. of the IEEE ICCV Workshops*, 2011, pp. 1154-1160.
[6]   T. Stoyanov, *et al., "Comparative Evaluation of Range Sensor Accuracy in Indoor Environments,"* in Proc. Eur. Conf. Mobile Robots, 2011, pp. 19-24.
[7]   W Budiharto, *et al*., "Multiple moving obstacles avoidance of service robot using stereo vision," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, 2011; 9 (3): 433-444.
[8]   M. Hamissi and K. Faez, "Real-Time Hand Gesture Recognition Based on the Depth Map for Human Robot Interaction," *International Journal of Electrical and Computer Engineering (IJECE)* , vol. 3(6), 2013. pp. 770-778.
[9]   D. Kuswardhana, *et al*., "The Instructional Thematic Game for Children with Mild Mental Retardation: For Enhancement of Left-Right Recognition Skill," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7(1), 2017. pp. 469-478.

[10] M. Z. A Bakar, *et al.*, *"Real-time rotation invariant hand tracking using 3D data,"* in Proc. IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 2014, pp. 490-495.

[11] M. Z. A Bakar, *et al.*, *"Finger application using K-Curvature method and Kinect sensor in real-time,"* in Proc. International Symposium on Technology Management and Emerging Technologies (ISTMET), Langkawi, Malaysia, 2015, pp. 218-222.

[12] E. Abu-Taieh, *"An Algorithm for Human Modeling in Information Technology Multimedia Using Human Biometrics Found in Golden Ratio, Vitruvian Man and Neufert,"* Fifth Int. Conference on e-Learning, Manama, Bahrain, 2015, pp. 65-73.

[13] M. Serrao, *et al.*, "Harmony as a convergence attractor that minimizes the energy expenditure and variability in physiological gait and the loss of harmony in cerebellar ataxia," *Clinical Biomechanics*, vol. 48, 2017, pp. 15-23,

[14] M. Iosa, *et al.*, "The Golden Ratio of Gait Harmony: Repetitive Proportions of Repetitive Gait Phases," *BioMed Research International,* vol. 2013, 2013, pp.1-7.

[15] Z. Zivkovic, *"Improved adaptive Gaussian mixture model for background subtraction,"* Proc. of the 17th Int. Conference on. Pattern Recognition, 2004, vol. 2.

[16] Z. Zivkovic, *et al.,* "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters,* vol. 27(7), 2006, pp. 773-780.

[17] Felzenszwalb *et al.*, "Distance transforms of sampled functions," Cornell University, 2004.

[18] G. Borgefors. "Distance transformations in digital images." *Computer vision, graphics, and image processing.* vol. 34(3), 1986, pp. 344-371.

[19] PrimeSense, OpenNI User Guide. OpenNI User Guide, 2011, http://doi.org/10.1007/SpringerReference_27988

[20] R. Samad *et al.,* "Elbow Flexion and Extension Rehabilitation Exercise System Using Marker-less Kinect-based Method," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7(3), 2017. pp. 1602-1610.

BIOGRAPHIES OF AUTHORS

Rosdiyana Samad was born in Selangor, Malaysia in 1980. She received the BEng. in Electrical Engineering from the Kolej Universiti Teknologi Tun Hussein Onn (KUiTTHO), Johor, Malaysia in 2002. In 2005 she received MSc. in Electrival Engineering from Universiti Sains Malaysia (USM) in Penang, Malaysia. She received PhD degree in Engineering (Intelligent Mechanical Systems Engineering) from Kagawa University, Kagawa, Japan in 2012. Currently she is a Senior Lecturer at Faculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang (UMP), Malaysia. Her current research interests include computer vision, image processing, pattern recognition, biomedical engineering, artificial intelligence and humancomputer interaction (HCI).

Law Wen Yan was graduated from Universiti Malaysia Pahang with Bachelor of Electrical Engineering (Electronic) in 2015. Currently he is an engineer and has been practicing it since 2015. He is experienced in firmware, microprocessor test, PLC circuit design and programming. He is currently involved in cryptography design and lives in Penang, Malaysia. He had achieved divisional award during his career in Intel.

Mahfuzah Mustafa is a senior lecturer in Faculty of Electrical & Electronics Engineering, Universiti Malaysia Pahang (UMP), Malaysia. She received her Diploma in Electronics Engineering from Universiti Teknologi Malaysia (UTM) Malaysia and received her BEng. (Hons) in Computer System & Communications from Universiti Putra Malaysia (UPM), Malaysia. She received her Master of Electrical Engineering from Universiti Tun Hussein Onn, Malaysia and received her PhD in Electrical Engineering from Universiti Teknologi MARA, Malaysia. Her current research interests include biomedical signals, medical imaging, signal processing and image processing.

Nor Rul Hasma Abdullah received her B.Eng (Hons) from Universiti Teknologi Malaysia in 2002; M.Eng from Kolej Universiti Tun Hussein Onn and Ph.D from Universiti Teknologi MARA in 2012. Her area of research interests are artificial intelligence, optimization techniques and power system stability. To date she is currently a senior lecturer and head of Power Engineering Reseach Group at Faculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang.

Dwi Pebrianti is a senior lecturer in Faculty of Electrical & Electronics Engineering, Universiti Malaysia Pahang (UMP), Malaysia since 2013. She received her bachelor degree in Electrical Engineering from Universitas Indonesia, Indonesia in 2001. She received her Master of Engineering in the Development of Gas Sensor from the Department of Engineering Synthesis, The University of Tokyo, Japan in 2006. Her PhD was received in 2011 from Chiba University, Japan in the field of Artificial System Science. She is a member of Institute of Electrical and Electronics Engineers (IEEE). Her main works are including image processing for robot navigation, control theory for robot navigation, automation, control system, robotics, wearable computer, artificial intelligence, nonlinear system and control, Unmanned aerial Vehicle, motion and dynamics control, motion tracking system and embedded system design.