

## Surveillance system with motion and face detection using histograms of oriented gradients

Ri Cerd Ng, Kian Ming Lim, Chin Poo Lee, Siti Fatimah Abdul Razak

Faculty of Information Science and Technology, Multimedia University, Malaysia

---

### Article Info

#### Article history:

Received Sep 25, 2018

Revised Nov 26, 2018

Accepted Dec 5, 2018

---

#### Keywords:

Face recognition

Histogram of oriented gradients

Motion detection

Surveillance system

---

### ABSTRACT

With the rapidly increasing crime rate in recent years, community safety issues aroused a wide concern among public community. Various security technologies had been invented and carried out, for example password door lock, alarm system, and closed-circuit televisions (CCTVs). Although the installation of CCTVs is common in most premises, they require extensive man power to manually monitor the videos. Moreover, the reliability of human operator greatly deteriorates when they are in fatigue condition. In view of this, our project aims to develop an automated computer vision based surveillance system. Unlike ordinary CCTV system that requires human operator to manually observe and detect intruder, a computer vision based surveillance system automatically monitor the security of premises and trigger actions once an intrusion is detected. Basically, it is a simple surveillance camera system that will be setup at the entrance of the house. The reliability is being enhanced by applying the motion detection and face recognition algorithm, using histogram of oriented gradients that could detect the existence of people at the main entrance and try to validate the user. Apart from recognizing the user, the propose system also support mobile interaction whereby user can monitor the camera, activate alarm, and even received notification when a stranger was being detected at the entrance of the house. By including such functionalities, proposed system had highly surpassed the existing surveillance system by not only support monitoring, but also try to recognize the people and inform the user at the exact moment when stranger detected, so that user could take immediate action about it, for example activating the alarm or report to police. The project was executed with expected outcome and objectives had been accomplished.

Copyright © 2019 Institute of Advanced Engineering and Science.  
All rights reserved.

---

### Corresponding Author:

Lim Kian Ming,

Faculty of Information Science and Technology,

Multimedia University,

Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia.

Email: kmlim@mmu.edu.my

---

## 1. INTRODUCTION

Crime is an illegal activity that is against the law and sometimes even threatens the life of the others. Crimes constantly happens around us, for instance burglary, robbery, murder, kidnapping, rape cases, theft cases and other types of crime. In this paper, a computer vision based surveillance system is proposed to detect burglary event. Burglary is an entry into a building or places with the purposes of committing an offence, and the most common offence is the theft cases. At the year of 2006, Malaysia burglary had 27,313 number of cases, a total increase of 11.64% from the year of 2005, these data were mainly focusing on unauthorized access to building or other premises [1].

In this paper, we propose an automated home-based burglary detection system empowered by computer vision algorithm. In the project, a surveillance camera is installed at the main entrance and

connected to a central processing system. Firstly, the camera will capture every single frame and the frame will be sent to the central processing system. Subsequently, the frame is analysed by some computer vision algorithms for motion detection and face detection. A motion detection algorithm determines whether there is any moving object in the housing area. The face detection algorithm, on the other hand, determines if a human being appears in the housing area. In the event that a face is detected, the face recognition algorithm will be activated. The face recognition algorithm aims to decide whether the face belongs to the residents or a stranger. If a stranger is detected, the system will send push notification to the pre-installed mobile phone application to inform the user. At the same time, the user can live stream the camera view in mobile phone to get the real time view of the scene when the stranger is detected. Besides, user also allow to trigger alarm from the mobile application, to warn the intruder about it.

## 2. MOTIVATION

A wide range of anti-burglary products have been developed to impede the occurrence of burglary cases. Examples of anti-burglary products are window or door alarm, deadbolt locks, keypad deadbolt locks, motion-activated lights, home alarm system, wireless surveillance cameras and so like. The most widely used anti-burglary product is the home alarm system and wireless surveillance camera. Home alarm system is usually delivered in a package, equipped with contact sensor, motion sensor, alarm system, and so on. The recent versions of home alarm system mostly include wireless surveillance camera in the package. Another interesting feature of the recent home alarm system is the mobile application that allows user to interact with the system through their phone. Which allow user to activate alarm, monitor the surveillance camera, check the sensor's status and others, via the mobile application.

Existing home alarm system and surveillance camera do provide sufficient protection and functionalities. Owning such a system, however, requires a high budget. From a preliminary online product survey, a set of such system motion detection, live streaming, recording functionality and mobile interaction functionalities costs around RM1,700 and more depends on the functionalities that user asked for.

Besides that, surveillance camera nowadays only provides the functionality to monitor the scene and video recording for further usage. However, in order to ensure the housing area is in safe condition, the user will need to monitor the area for 24 hours per day, or else they might miss the moment when the thief is trying to break into the house and unable to take the immediate actions about it. To overcome this issue, the security companies do offer a monitoring service to the user, whereby there will be professional to help the user monitor their housing area for 24 hours, and inform the user immediately once suspect burglary is detected. To subscribe for the service, user will need to pay monthly fees to the security company. The main issues of this solution is about the privacy of the user. The monitoring service basically allows the third party to monitor not only the housing area of the user, but also the user's daily activities. In view of this, a surveillance system that requires low cost components and automatically monitor the housing area is developed.

## 3. SUVEILLANCE SYSTEM WITH MOTION AND FACE DETECTION USING HISTOGRAMS OF ORIENTED GRADIENTS

This section describes the system flow and functionalities of the proposed surveillance system.

### 3.1. System Architecture

The proposed automated computer vision based surveillance system is divided into two main parts that run separately on different operating environments and devices. The main subsystem runs on a Linux based operating system and is coded in Python3. The second subsystem is an Android application coded in Java language. Both subsystems are connected through local TCP handshake protocol and local HTTP web server. The overall process model and system architecture are illustrated in Figure 1 and Figure 2. The use case diagram in Figure 3 depicts the functionalities of the proposed system. Following by the general activity diagram of the overall system at Figure 4 and separating the functionalities most detailed activity diagram as showed in Figure 5 the activity of motion sensing and face recognition. Figure 6 refer to the activity diagram for live streaming and alarm trigger from mobile application, the last Figure 7 represent the activity diagram of live stream function from the system.

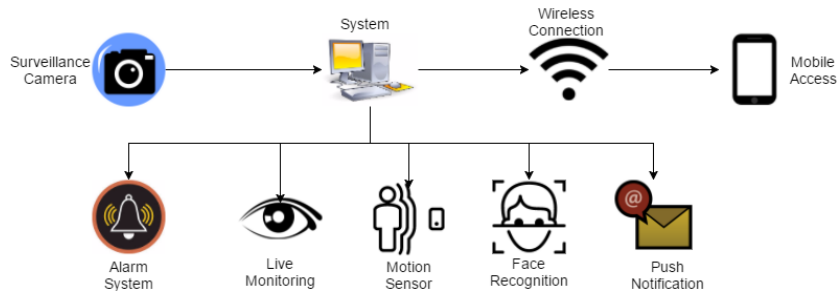


Figure 1. Process model

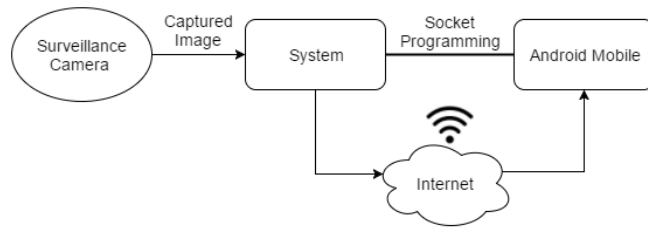


Figure 2. System architecture

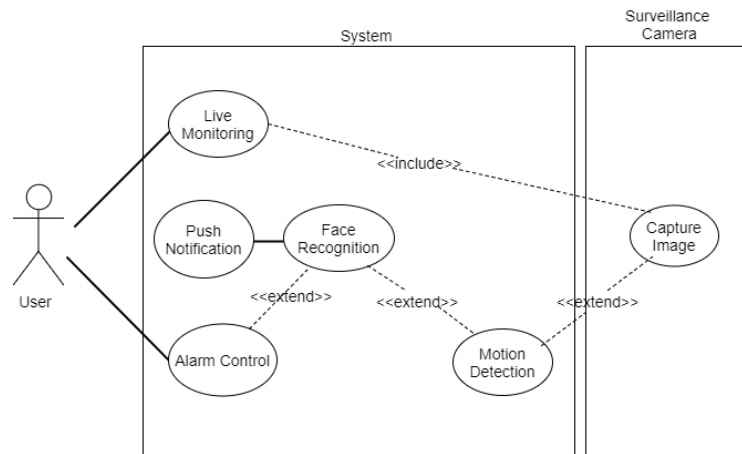


Figure 3. Use-case diagram

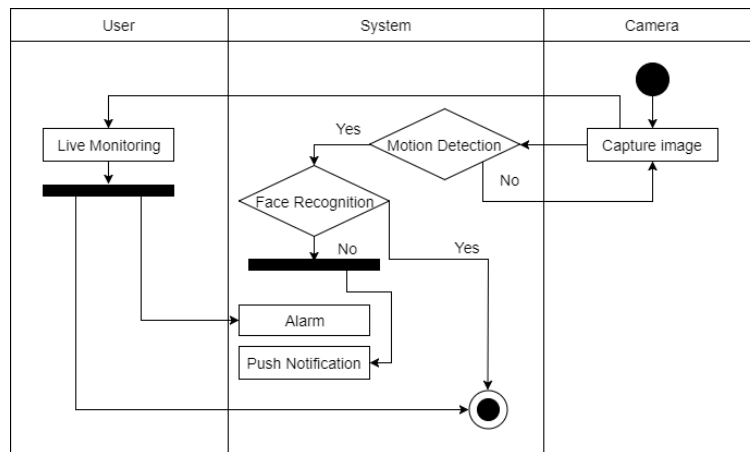


Figure 4. General activity diagram

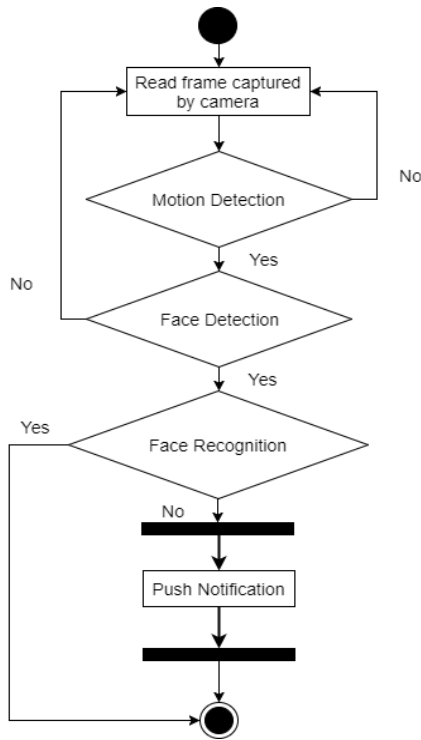


Figure 5. Activity diagram on motion sensor and face recognition from system

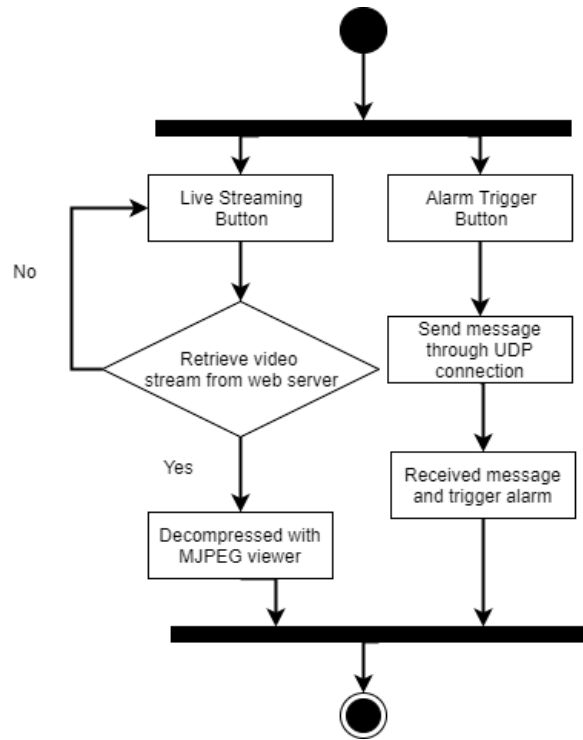


Figure 6. Activity diagram on live streaming and alarm trigger from android application

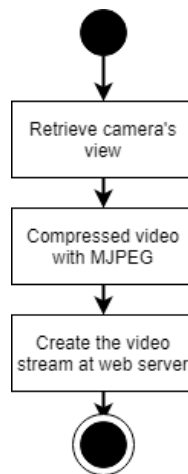


Figure 7. Activity diagram on live stream from system

**3.2. System Functionalities**

Before using the application, user will need to setup the system and the hardware to maximise the usability and functionalities of the system. The camera is recommended to be setup at the entrance of the house. It is also advisable to setup the camera at an angle where the camera is able to capture the entire face of people for better face recognition. The following subsections explain the working of functionalities and the interaction process of the system.

**3.3. Image Capturing and Motion Detection**

The system starts by acquiring images from the camera and use it for motion detection. An open source library named OpenCV (Open Computer Vision) is used to acquire the image frames from camera to the system and allowing the system to perform further processing with the images.

After the images are ready to be used by the system, next the system executes motion detection to detect movement in the camera’s view. This function is being done by applying background subtraction that calculates the difference between subsequent frames as shown in Figure 8. An image difference that surpasses the thresholding value indicates the possible existence of motion in the view. Once motion is detected, the specific frame of image will be sent to the next process which is the face detection process.

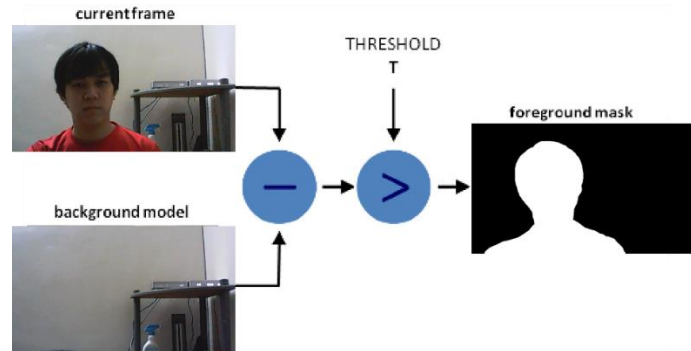


Figure 8. Example of background subtraction for motion detection

**3.4. Face Detection**

This process is mainly to detect human faces in a single image, to be used in the face recognition process. Face detection will only be performed after the motion is detected, and it will use the frame from the motion detection process. This process is done by combining several steps, namely face detection, posing and face projection.

The process starts with face detection using the method called Histogram of Oriented Gradients (HOG), which works good in either bright or dark images. In order to use the HOG method, the image will need to be converted into grayscale, then it will compare the intensity of the darkness of a single pixel with its neighbouring pixels. The HOG will apply on every single pixel on the image, and lastly all the pixel will be replaced by an arrow that shows the gradient flow from light to dark. To prevent the HOG from giving too detail result, the image is partitioned into square regions with  $16 \times 16$  pixels each. The system then counts how many gradient points in each major direction and replaces that square with the arrow directions that are the strongest. After encoding the detected face in HOG form, next we compare the detected face HOG with another HOG that were generated from lots of human face images, as depicted in Figure 9, to determine whether the detected face is a human face [4].



Figure 9. HOG face pattern generated from lots of face images

The process continues with posing and projecting faces, in order to deal with faces that are in different directions or angles. To done the following task, face landmark estimation algorithm is being apply to this situation, that a machine learning algorithm that able to detect 68 specific human face landmarks on a face image as shown in Figure 10. After the face landmark detection, we can know the position of face feature in an image, so that we can easily rotate, scale and shear the image to make sure the face features are in the centre of the image, to make the next process a lot easier to be done [5].

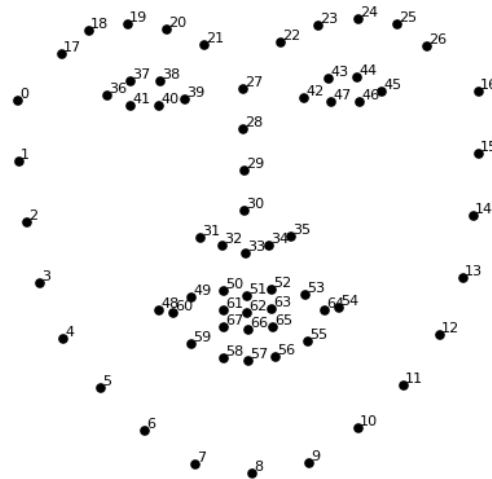


Figure 10. The 68 specific human face landmarks

### 3.5. Face Recognition

After detecting face and centralising the face feature in a single image, the face recognition process is conducted to recognise the identity either a user or a stranger. In the basic face recognition method, the unknown face is compared with the tagged faces, which is time consuming. In this project, a face recognition method is proposed to overcome the time-consuming problem for the basic face recognition approach. The face recognition method includes three main steps namely face encoding, feature extraction and classification.

In the face encoding and feature extraction step, deep learning is implemented in this step to determine which part of the face is important to be measured for the face recognition process. The deep convolutions neural network is being trained by loading a training face image of a known person, load another picture of the same known person, and load a picture of a totally different person, then the neural network tweak accordingly based to the result. After lots of repeating training, the neural network learns to reliably generate 128 measurements for each person. To save the works and time of training the neural network, in this project we implemented one of the trained network from Open Face to get the 128 measurements for our images [3], [6], as depicted in Figure 11.

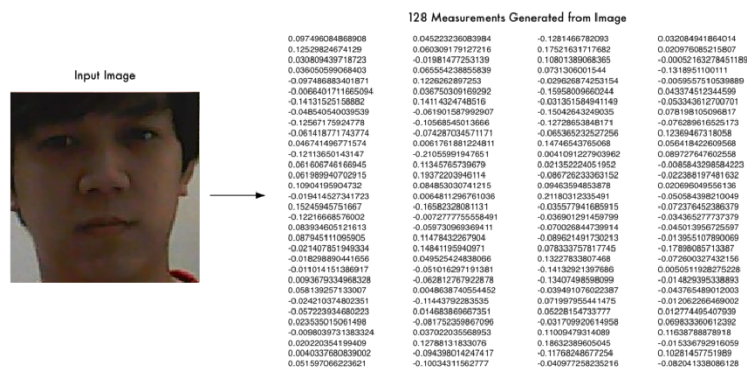


Figure 11. Example of 128 measurements extracted from human face

After the extraction of 128 measurements from the face, then the process is continuing with the last classification step. For the classification method, we used the linear Support Vector Machine (SVM) classifier to compare the 128 measurements of an unknown face with the tagged face, and based on the result to classify the specific face as stranger or user [2].

### 3.6. Trigger Notification

Once a stranger is detected by the face recognition process, the system will trigger a notification to the mobile application. The process is done by using the Google Firebase Cloud Messaging (FCM) application program interface (API) which support notification trigger function at the Android application. In order to run the FCM notification trigger function, a project need to be created at the Firebase console, and by using the server and client key generated by Firebase, apply it to the FCM API and the notification are easily be done. After the user received the notification, they are allowed to remotely communicate with the system via the application, for instance, triggers the alarm and live streams the home scenes function.

### 3.7. Alarm Trigger

The alarm trigger function mainly depends on the Transmission Control Protocol (TCP) connection as a medium between two systems, and sending messages to trigger some functions. In alarm trigger, once the user activates the function from the mobile application, the application will send a message to the system. When the system receives this specific message, it will trigger the alarm function. All the messages sent through the TCP connection are set to be different, to differentiate between the trigger of functions.

### 3.8. Live Streaming

In the live streaming part, the camera's view is compressed into a single JPEG image using Motion JPEG (MJPEG). After compressing the video, next is to reside the video on the local web server using Python web framework API called Flask that enables the streaming of the camera's view on the web server. Basically, the live streaming will first stream to the web server so user can directly view from the web server, or on the other side they could also view it through the mobile application. The mobile application is designed with MJPEG viewer, which could allow the mobile application to convert the JPEG images into video. The mobile application will subsequently retrieve the MJPEG video stream from the web server and play it at the mobile application.

## 4. RESULTS AND ANALYSIS

Testing ensures the level of performance, stability, and acceptance, thus brings significant improvements and refinements to the system. User Acceptance Test (UAT), as presented in Table 1, was conducted for the system to ensure the system functionalities functioning in the expected manner and to make sure the system is solving the needs and fulfilled the requirements.

Table 1. Results of UAT

Number	Test Cases	Expected Output	Accept/Reject
1.0	Starting TCP server	Server successfully created and listening for client.	Accept
1.1	TCP client try to connect the server	Client connected to the server and ready to pass data.	Accept
1.2	Server receive message from client	Server received the message from client and perform specific function.	Accept
2.0	Retrieving camera's view	Camera's view retrieved and pass for motion detection.	Accept
2.1	Perform motion detection	Threshold view being displayed and able to detect motion from the frame.	Accept
2.2	Face detection	Detect face from motion detected frame, if face detected bounding box will be created at the face in the display window.	Accept
2.3	Face recognition	Classify as user or intruder.	Accept
2.4	Trigger notification	Push notification received at the mobile device	Accept
3.0	Trigger alarm in mobile application	System play alarm sound.	Accept
4.0	Start live streaming in mobile application	Camera's view being displayed at the mobile application.	Accept
4.1	Stop live streaming	Camera's view being disable at the mobile application.	Accept

## 5. CONCLUSION

As a conclusion, the proposed system has achieved its objective in terms of enhancing surveillance camera by applying computer vision algorithms. Several techniques such as motion detection, face detection, and face recognition algorithm are implemented in the system. Firstly, the system performs motion detection using background subtraction method between subsequent video frames. Secondly, the frames with motion will be passed to the face detection phase to locate the face region. In this phase, the HOG method is used as the face descriptor. Thirdly, face recognition is performed on the face region. In this paper, a deep convolutional neural network architecture is used to recognize the identity of the face.


## ACKNOWLEDGEMENTS

Research reported in this paper was supported by Multimedia University Mini Fund, Grant No. MMUI/170032 and MMUI\180182. The research was also sponsored by MMU CAPEX Fund MMUI/CAPEX 180011. The Quadro P6000 used for this research was donated by the NVIDIA Corporation.

## REFERENCES

- [1] Knoema. (2017). Malaysia Burglary, Car Theft and Housebreaking Burglary, 2003-2014 - knoema.com. [online] Available at: <https://knoema.com/atlas/Malaysia/topics/Crime-Statistics/Burglary-Car-Theft-and-Housebreaking/Burglary> [Accessed 10 Sep. 2017].
- [2] Geitgey, A. (2017). How to do Modern Face Recognition with Deep Learning. [online] iamwire. Available at: <http://www.iamwire.com/2017/03/modern-face-recognition-deep-learning/149697> [Accessed 10 Sep. 2017].
- [3] B. Amos, B. Ludwiczuk, M. Satyanarayanan, "Openface: A general-purpose face recognition library with mobile applications," CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.
- [4] Watanabe, T., Ito, S. and Yokoi, K. (2010). Co-occurrence Histograms of Oriented Gradients for Human Detection. IPSJ Transactions on Computer Vision and Applications, 2, pp.39-47.
- [5] V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees."
- [6] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering."

## BIOGRAPHIES OF AUTHORS

	<p>Ri Cerd Ng completed his B.IT (Hons) in Artificial Intelligence in the Multimedia University, Malaysia. After completion of his study, he joined ViTrox Technologies, Malaysia as a Research and Development Engineer that mainly focus on development of machine vision based software and hardware solution for semiconductor field.</p>
	<p>Kian Ming Lim received B.IT (Hons) in Information Systems Engineering, Master of Engineering Science (MEngSc) and Ph.D. (I.T.) degrees from Multimedia University. He is currently a Lecturer with the Faculty of Information Science and Technology, Multimedia University. His research interests includes machine learning, deep learning, computer vision and pattern recognition.</p>
	<p>Chin Poo Lee is a Senior Lecturer in the Faculty of Information Science and Technology at Multimedia University, Malaysia. She completed her Masters of Science and Ph.D. in Information Technology in the area of abnormal behaviour detection and gait recognition. She is a certified Professional Technologist since 2018 and currently the Deputy Director of ADEPT (Office of Academic Development for Excellence in Programmes and Teaching), Chairperson of the Admission and Credit Transfer Committee, a senior researcher of a few Mini Fund projects, and Project Leader of an external grant project funded by MOHE. Her research interests include action recognition, computer vision, gait recognition, and deep learning.</p>
	<p>Siti Fatimah Abdul Razak received her B.Sc (Hons) with education where she majors in Mathematics and Information Technology and Master of Information technology majoring in Science and System Management from the National University of Malaysia in 2004. She completed her postgraduate studies in Information Technology from Multimedia University. She is currently a lecturer in Faculty of Information Science and Technology, Multimedia University. Her research interest includes rule mining, information systems development and educational technology.</p>