Real-Time Detection of Suspicious Human Movement

Sugla Vinaayagan Rajenderan[a], Ka Fei, Thang[b]

School of Engineering, Asia Pacific University of Technology & Innovation
Technology Park Malaysia, Bukit Jalil, 57000 Kuala Lumpur, Malaysia
[a]suglavinaayagan@gmail.com , [b]ka.fei@apu.edu.my

ABSTRACT

Crime is everywhere and it could be argued that we are in one of the most crime eras in human history. Crimes like theft, violence against people and property damage are some of the rising crimes in university. On the other hand, investigation is held to find the person who is responsible for the theft, only after the crime occurred, by using existing surveillance system. It acts as 'post-mortem' tools. Therefore, in this paper, an appropriate algorithm for autonomous suspicious human-movement detection from surveillance videos is proposed. First, proposed system extract human-movement information by detecting and tracking people in real-time using Gaussian Mixture Model (GMM). Morphological operations aid the detection of human-movement by eliminating noises. Features extracted from human-movement are then sent for post-processing in order to recognise whether the detected activity is suspicious or not. Examples of suspicious movements are loitering and hanging or looking around in the area of interest for longer time period. The framework used for recognising suspicious activities is called Grammar-based approach. This approach is effective in detecting suspicious activities with serial of events for a longer time period. By using recorded videos with people mimicking those suspicious activities, several experiments have been performed and results presented in this paper. The experimental results presented here demonstrate the outstanding performance and low computational complexity.

Keywords: Human-Movement Detection, Gaussian Mixture Model, Grammar-Based Approach

1 INTRODUCTION

In recent years, act of violence in the world is rapidly growing. Humankind has sought for safety and security since the beginning of our existence. Hence, demand for security system is growing at a very fast pace to secure the safety of humankind. For an instance, after the Boston Marathon bombings, the ruling party of United States of America increased the use of surveillance camera to maximise the security in the country (Huffington Post, 2013).

Therefore, surveillance camera system prevails as a security system in commercial, military application and now in educational institute like university. Before advancement of technology, surveillance cameras were mounted and they provide visual information which can be viewed as live videos, or it can be played back for future references. These surveillance cameras were not able to prevent any attacks from happening but they can provide adequate information of the incident which is very helpful for the authorities for post-incident analysis. For an instance, for Boston terrorist attack, The Wall Street Journal (2013) praises the practical value of these cameras in speedy identification of the Boston Marathon terrorist.

Although live footage is presented, to find available human resources who can monitor the video for hours are not practical. Most researchers pointed out that the current surveillance systems are sufficient as 'post-mortem' tools. The reason for the statement is that security agencies or appropriate authorities use the footage for investigation and legal purposes after that incident. Moreover, surveillance cameras play the role of deterrence effect. Deterrence effect simply means a mere threat for the people so that they can modify their behaviour when they notice the presence of such system. In considering crime prevention

ISBN: 978-1-941968-02-4 ©2014 SDIWC

systems, this deterrence effect could be achieved in some areas. However, Lepon & Popkin (2007) stated that as people not aware of the system, the crime will increase; hence deterrence effect is not a solution for crime prevention. Moreover, Lepon & Popkin (2007) also shows a study whereby only 10% of surveillance cameras in London watched 24/7. Plus, while monitoring the video feeds, the 82% of observers have to do simultaneous duties such as checking in visitors, attends calls and etc. Consequently, manual monitoring of a place may be prone to error during human supervision.

Hence, in this paper an appropriate algorithm for suspicious human-movement detection from surveillance videos is proposed. Such a system has previously been developed but it is still challenging because there are many issues that need to be addressed. One of the questions needed to be addressed is how 'intelligent' is the system to understand the human behaviour in university. Therefore, this research will focus on better understanding of human-movement in university predominantly suspicious behaviour, in order to recognise the intruder using human movement-detection algorithm. Hence real-time detection of suspicious human movement will overcome the shortcomings and provide perfect automated surveillance system in educational institutions as in this case universities.

2 LITERATURE REVIEW

2.1 Video-Processing Methodologies

The typical architecture of video surveillance system consist of video cameras, video processing unit (VPU), network, visualisation centre and video database and retrieval tools . The front-end of the system starts with camera as it captures video. The video is transmitted to the VPU through network as it enters middle-end of the system. This unit is solely software-based system where data-intensive video analytics (VA) algorithm such as background modelling will be developed.

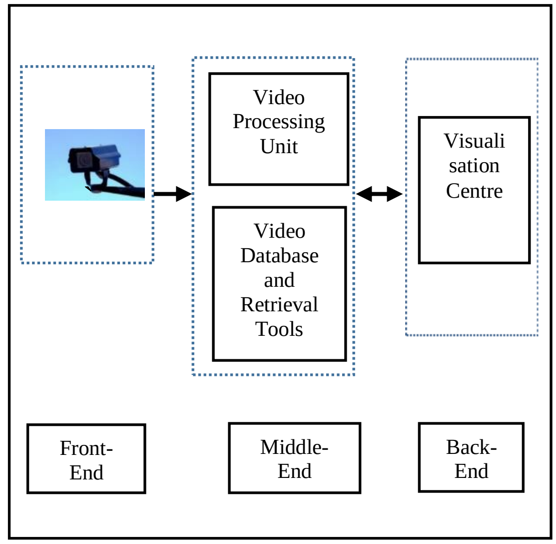


Figure 1: Architecture of video surveillance system

The middle-end system equipped with video storage and retrieval tools. This will aid the VA for processing algorithms such as storing the related content of a video for further processing. Useful information which is gathered in VPU is transmitted to visualisation centre whereby a human monitor may response to the alarm resulted from an intruder. This is the part where the real interaction occurs between human and the system and it is known as the back-end. (Chan *et al.*, 2013). The architecture of the video-surveillance system is illustrated in Figure 1.

From the architecture presented, VPU is the most significant part in the whole video-surveillance system. There are plenty of video related processes that can be performed in VPU. However, Guo *et al.* (2013) emphasis that, in moving object-detection, background subtraction is a difficult task because the background is not always in a static position. Moreover, they explained the significance of noise removal in a video or image processing in their project. An effective noise reduction leads to an efficient video process, (Guo *et al.*, 2013). Since suspicious human-motion will be detected in this paper, a review on background

ISBN: 978-1-941968-02-4 ©2014 SDIWC

subtraction is presented for better understanding.

2.2 Background Subtraction

Background subtraction or background modelling is a process to extract a foreground object from a video for further analysis. The logic behind background subtraction in detecting moving object is to differentiate current frame with a reference frame. The reference image is commonly known as background image. Few techniques have been proposed in recent years to perform background subtraction.

Lee (2005) proposed adaptive Gaussian Mixture modelling for modelling static temporal distributions of pixels in video. Unlike the normal Gaussian modelling for background subtraction, it has adaptive learning rate calculation for each Gaussian at every frame. This lead to a stabilise modelling with better convergence rate. However, it has certain drawbacks; for an instance it is sensitive to gradual changes of illumination. These sudden changes in illumination will result in inefficient background subtraction.

Gaussian Mixture Model has attracted many researchers in developing background subtraction algorithm. However, Guo *et al.* (2013), recently uses codebook (CB) as background subtraction in video processing especially in extracting moving object. They use multilayer codebook-based background subtraction to eliminate static backgrounds. This will increase the efficiency of video processing. Moreover, it also can identify foregrounds, shadows and highlights by classifying the pixels of a video.

Figure 2 (Guo *et al.*, 2013) shows the result obtained by (a) Original Image (b) Conventional codebook method (c) using Fast Codebook without light (d) Final result of Fast Codebook method with proper lighting. Hence, this method successfully performs background subtraction. They also mentioned that, compared to GMM, this codebook-based background subtraction can provide better performance.

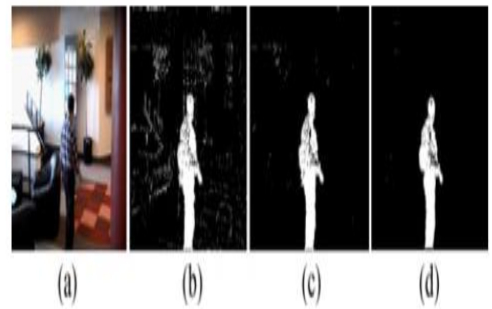


Figure 2: Background Subtraction result (Guo et al., 2013)

2.3 Human Movement Detection Methods

Human movement is a wider term that covers a lot of categories of activities, which require different means of detection. As mentioned by Moeslund, Hilton & Kruger (2006), when compared with other field of research in computer vision, the field of action and activity recognition and representation is relatively old and challenging. Hence, a lot research has been done to the numerous type of human movement. There are many researchers researched about human behaviour in different perspectives. In the context of human movement, the approaches can be divided into three main categories. Those are scene interpretation approach, holistic approach and grammar-based approach. Each approach is explained in the respective headings.

A) Scene Interpretation

Scene interpretation approaches utilise the whole scenery without making any model of humans or particular objects. This approach considers the view of the camera as a whole and learns the activities simply by monitoring the motion of objects. The significant of this approach is, it does not need the identity of the person who is the reason for the motion of the object. In order to detect drowning in a

ISBN: 978-1-941968-02-4 ©2014 SDIWC

swimming pool, Eng *et al.* (2003) proposed a surveillance system in pool. They use the scene as a tool to interpret and detect drowning events rather than observe swimmers' movements. Submersion index, activity index, splash index, speed and posture are the five significant parameters to describe the swimmers action. Although they used the scene for developing the algorithm, tracking the swimmer is essential. Markov Random Field framework is used to enhance the tracking capability system. This framework has the ability to eliminate partial occlusion, which is usually occurred in conventional tracking system. However, the main drawback of the project is this surveillance system cannot be used in other domain. The reason is the parameters used for scene interpretation is constrained in swimming pool (water) where it can detect water crisis only.

B) Holistic Approach

Holistic approach is completely different from scene interpretation in a way that this approach uses body dynamics. It can be the either whole body movement or movement of any body parts. Hsieh *et al.* (2008) full utilise the whole human body to determine human posture. They used mixture Gaussian model to reconstruct the background from the video since the camera is in stationary position; followed by video de-noising, which is done by using morphological operations.

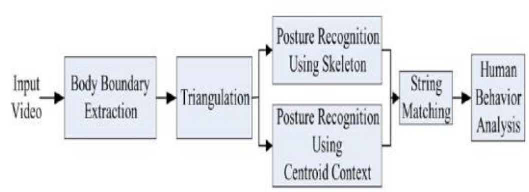


Figure 3: Flowchart of Human-motion recognition (Hsieh *et al*., 2008)

The process of the system is well illustrated in the flowchart in Figure 3 (Hsieh *et al.* 2008). There are predefined control points extracted along human posture contour in advance in this algorithm. High curvature points of boundary points which are along the posture are extracted to find the posture of human. Instead of using silhouette based approach which used conventionally, they proposed Delaunay Triangulation algorithm that is robust and efficient. This triangulation algorithm divides human posture into triangular meshes. As a result, this algorithm will extract the centroid context as well as skeletal features directly.

C) Grammar-based Approach

Turaga *et al.* (2008) stated that several simple events that occurred sequentially for a long duration are denoted as "activities". Simple events which are classified in a system, are not adequate to understand the activity which occurs for long time, as done by other two approaches. Grammar based approach will be used in complex events that has serial of events. In this approach, a set of rules which describes human behaviour is applied in order to detect suspicious human activity. If a person's movement tallied with predefined human movement, the system will notify the developer. However, this approach can only be applicable if the suspicious human movement is known. Moreover, since serial events extracted and predefined, Li *et al.* (2013) insist that suitable machine learning algorithm or classifier is needed to categorise the events in order to recognise the human-movement.

3 CONCEPT DESIGN AND RESEARCH METHODOLOGY

3.1 Proposed Methodology

An intelligent system comprises of suspicious human-movement detection algorithm and is integrated with alert system as illustrated in Figure 4. At first, real world video is fed into the system by using a webcam. Webcam is implemented on the ceiling to work as surveillance closed-circuit television (CCTV) to get video as it is from surveillance camera. The video from the webcam undergoes pre-processing. There are two main sub-process in this pre-process, background subtraction and noise removal. Human-motion is extracted using an algorithm called Gaussian Mixture Model (GMM). GMM is the algorithm for background subtraction in detecting motion object. However, GMM is sensitive to illumination changes. Hence, an effective noise removal is needed in order to effectively track human-motion in the video. Morphological operation will aid in the human-tracking by eliminating unwanted pixels eventually extracted pixel information about human movement only.

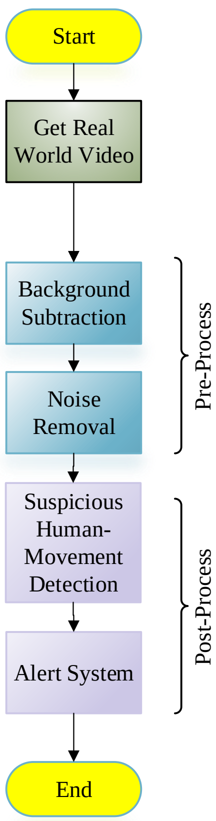


Figure 4: Flowchart of proposed system

In this paper, morphological operation is also used to reduce errant movement when human in motion. Errant movement in human motion is merely minute movement of hand or leg movement. Generally, in human locomotion each of the body parts move in different angle and position. Consequently, there are chances for the GMM identified a single human movement into several movements due to the uneven movement. This phenomenon is known as errant movement. Morphological operation is used to rectify the aforementioned problem with proper algorithms. Extracted information which is movement of human is transmitted to the post-processing process where suspicious human-movement detection will occur. The system determines whether the detected human movement is suspicious or not. This method is called grammar-based approach as there will be several suspicious activities predefined in the algorithm, which will aid in the suspicious human-movement recognition. In this paper, two types of suspicious movement are predefined, which are “loitering” and “looking-around”. If the system recognised the human-movement as suspicious the alert system will alert the security officer. The alert is in the form of sound as well as text on the screen of the video stating the type of suspicious movement recognised.

3.2 Pre-Processing Techniques

A) Gaussian Mixture Model

GMM represents weighted sum of Gaussian component densities. Hence, it is known as parametric probability density function, and commonly used as a parametric model of the probability distribution of continuous measurements in a biometric system. Image is matrix which each element is a pixel. Hence, in this case, it model the pixel related spectral features in an image. Value of pixel is a number that shows intensity or a colour of the image. Suppose X is the random variable that takes the values of pixel, whose probability density function represented in the following equation:-

$$P(X_t) = \sum_{k=1}^{K} \omega_{i,t} \eta(x|\mu_{i,t}, (\sigma_{i,t})^2) \qquad (1)$$

Where $P(X_t)$ are component of Gaussian densities and $\omega_{i,t}$is weight related to Gaussian

ISBN: 978-1-941968-02-4 ©2014 SDIWC

with the mean, μ and standard deviation, $(\sigma_{i,t})^2$. This Gaussian weight must be more than zero, for the number of Gaussian distribution, *k*=1, 2…*K*. On the other hand, Gaussian probability density function denoted as η is can be further expanded to become:-

$$\eta(x|\mu_{i,t},(\sigma_{i,t})^2) = \frac{1}{(2\pi)^{n/2}|(\sigma_{i,t})^2|^{1/2}} e^{\frac{-(x-\mu_{i,t})^2}{2(\sigma_{i,t})^2}} \quad (2)$$

Hence, each pixel of an image is represented by a mixture of *K* Gaussians. First step in background subtraction is initialisation of significant parameters. Number of Gaussian distribution, the weight related with time and mean as well as the standard deviation are the significant parameters to be initialised. Number of Gaussian distributions, or in computer algorithm is better known as number of Gaussian modes preferably set to be 3 due to suitable memory and computational power as mentioned by Bouwmans, Baf & Vachon, (2009). Other parameters are can be determined by systematic trial in this paper.

After the initialisation, foreground detection is done followed by updating of parameters. The basic idea of background subtraction is to subtract current frame with predecessor frame to find foreground object. Therefore, when new frames incomes at times *t*+1, a match test is made for each pixel to detect foreground object. In order to detect the foreground object, initially, background pixel corresponds to a high weight with a weak variance is extracted using GMM. Then, for new frames, following equation is aid the matching process where a pixel matches a Gaussian distribution if:-

$$\sqrt{\left[\left(X_{t+1}-\mu_{i,t}\right)^T \Sigma_{i,t}^{-1} \cdot \left(X_{t+1}-\mu_{i,t}\right)\right]} < k\sigma_{i,t} \quad (3)$$

Where *k* is threshold of an image. If a match is found with one of the Gaussian distribution, then it classified as background pixel (image). On the other hand, if no matches are found due to pixel changes from foreground movement, the pixel is classified as foreground. Accordingly, for each new frame the parameters gets update and matching process will be done for each pixel. Consequently, foreground detection can be made.

B) Morphological Operation

In order to accomplish objective of analysis and recognition image, such as reduction image data by keeping basic shape character as well as elimination dissertation structure, a mathematical operation is created to ease the analysis of image. It is known as mathematical morphology or simply morphology. It uses structuring element to measure and distil corresponding shape of an image. In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbours. Thus, selection of size and shape of the structuring element is very vital in construction a suitable morphological operation which is sensitive to specific shapes in the input images.

There are four basic morphological operations; dilation, erosion, opening and closing. Each operation has its trait respectively in binary and grayscale image processing. Dilation is a process adding pixels to the boundaries of objects in an image. It is performed by laying the structuring element denoted as B on an image denoted as A and sliding it across the image in a manner similar to convolution. The notation for the dilation process is as follows:-

$$A \oplus B = \{z|(\hat{B})_z \cap A \neq \emptyset\} \quad (4)$$

Mainly, there are two types of scenario faced in the dilation process. Those are:

1. If the origin of the structuring element overlaps with a ‘white’ pixel (grayscale image), there will be no change, eventually move to next pixel
2. If the origin of the structuring element overlaps with a ‘black’ pixel (grayscale image) it will fill in with

maximum value of all the pixels in the input pixel's neighbourhood.

On the other hand, erosion removes pixels on object boundaries. The number of pixels added or removed from the objects in an image depends on the size and shape of the structuring element used to process the image. In erosion, structuring element, B is shifted by displacement, z. The process of erosion is defined as follows where A is the input image:-

$$A \ominus B = \{z | (B)_z \subseteq A\} \quad (5)$$

Same as dilation there are two types of scenario faced in this erosion process. Those are:-

1. If the origin of the structuring element overlaps with a 'white' or 'black' pixel (grayscale image), there will be no change, eventually move to next pixel
2. If the origin of the structuring element completely inside, then the output pixel is minimum value of all the pixels in the input pixel's neighbourhood.

Morphological open operation is erosion followed by dilation. It generally soothes the contour of an object, breaks narrow isthmuses, and eliminates thin protrusions. Noises which are background of an image can be successfully eliminated in the erosion stage of opening because in the case all noise components are physically smaller than the structuring element. However, new 'gaps' or elimination of foreground image performed in the erosion process. In order to counter undesirable effect, a dilation process is performed on the opening.

In contrary to opening, morphological close operation is dilation followed by erosion. It also tends to smooth sections of contours but, as opposed to opening, it generally fuses narrow breaks and long thin gulfs, eliminates small holes, and fills gaps in the contour.

3.3 Post-Processing Techniques

After the pre-processing, the videos will undergo post-processing to determine the movement is suspicious or not. Since, the proposed system mimics human role as security personnel, hence it will be designed as it detects suspicious-movement that can be easily noticeable by human in short period of time. A study done by Stedmon *et al.* (2008) concludes several suspicious movements that can be noticed by human. Stedmon *et al.* (2008) stated that, looking around the premises, staring or looking at something for quite some time and loitering are the top most suspicious human-movement in universities. In this thesis loitering and looking around in the place for longer period of time are used as suspicious human movement. All these activities are classified in this system in order to detect suspicious movement in the university.

From the literature review there are three different human-movement techniques used by researchers. Those are scene interpretation, holistic approach and grammar based approach. First, the scene interpretation is not suitable for this system. Basically, it detects the impact from the human-movement in order to identify the human movement. In this case, the impact of each human-movement is resulted in crime. The main aim of this smart security system is to prevent any prominent crime before it happens. Hence, scene interpretation is not suitable for this paper. The second technique is holistic approach where this approach utilises human body to determine human-movement. This approach will be used in this paper. The reason is the suspicious human-movement defined by Stedmon *et al.* (2008) can be determined by monitoring human body parts. For an instance for looking around, head movement can be monitored. Hence by extracting the head movement, suspicious movement can be determined. However, each of the suspicious movements has a series of events, such as for looking around, the human have to move to different positions. Holistic approach does not have the ability to detect those different movements for a time period. Hence, third approach which is grammar based approach is well-suited in this paper. In this approach, features are extracted holistically and several

ISBN: 978-1-941968-02-4 ©2014 SDIWC

rules are predefined in order to detect the series of events in a suspicious human movement.

4 FEATURE CALCULATION

Table 1: Feature extracted from the video

Feature	Definition	Algorithm
Position	Location of the person in the context	Centroids of the person which extracted using GMM and bounding box.
Direction	Direction of the person to determine which way the person is moving, either towards the entrance or away from entrance	Current location of a person compared with previous location of the person. The comparison is done at each second of the video.

GMM is used in pre-processing to track human-movement in the video. After tracking the human-movement in the video, motion features needed to be calculated and send to the post-processing for the suspicious human-movement detection. For extracting important features about human movement, bounding box is created at each frame as GMM tracks human movement; a rectangular box will enclose the whole human. In digital image processing, the bounding box is merely the coordinates of the rectangular border that fully encloses a digital image. Hence, by enclosing human in a box, coordinates of the person can be calculated by obtaining the centroid of the box.

By extracting the information about human location, direction of human motion can be determined. This determination is further calculated in pre-processing. Table 1 shows the features that are extracting in pre-process and their explanations.

5 DETECTION OF SUSPICIOUS HUMAN MOVEMENT

5.1 Loitering

Table 2: Scenarios for the loitering

Scenario	Description
Person in ROI <30seconds	Person is at the entrance and timer is counting continuously if the person is remain at entrance
Person in ROI>30seconds	Person is identified as loitering
Person in and out from ROI	When the person is in the ROI, timer is start. However when he leaves the ROI, timer will reset.

Candamo et.al (2011) defined loitering as “the presence of an individual in an area for a period of time longer”. Moreover Stedmon et.al (2008) stated that it is the top most suspicious behaviour that likely to result in criminal act in not only educational institutes, but in public places like shopping centre and train station. However, one of the major questions is developing algorithm for detecting loitering is the place where the person is loitering. In this paper the region of interest (ROI) is the entrance. Consequently, loitering will be useful in detecting potential sneak in which greatly occurs at the entrance. Time is another important parameter in loitering since Candamo et.al (2011) defined loitering with respect to time. 30seconds of loitering time is selected for the experiment. The algorithms for the detecting loitering in the area are developed by considering three types of scenarios. Table 2 shows the three scenarios and respective descriptions.

5.2 Looking Around

In most of the cases, loitering and looking around are considered as same kind of human behaviour because both have time as threshold.

However, Stedmon et.al (2008) distinguishes looking around from loitering by the movement of the person. They also mentioned, looking around is top five human movement likely to be regarded as suspicious. Loitering is focused at a particular place where a person is standing in that place without any obvious movement. Mostly, loitering related with criminal activity like sneaking. However, looking around is hanging around the place to study the nature of the place. Stedmon et.al (2008) stated, person who are looking around are planning something suspicious for an instance measuring tangible cost materials in the area. Hence, it is important to identify that kind of movement in order to prevent any theft of tangible cost materials. For this type of behaviour, since the area is wide 60 seconds is selected in this thesis. Three types of scenarios are considered in developing this suspicious human-movement. Table 3 shows the three scenarios and respective descriptions.

Table 3: Scenarios for looking around

Scenario	Description
Person in the context <60seconds	Person is in the area and timer is counting
Person in the context >60seconds	Person is in the area for more than 60 seconds
Person in ROI (overlaps with loitering)	If the person is ROI, timer for this particular algorithm is continuously counting, unless person is detected as loitering first.

6 RESULTS AND ANALYSIS

The detection of suspicious human movement is challenging due to several reasons. First, most activities of interest are of high complexity, which becomes an issue in the clutter scene. Another issue is the inadequacy of professional and challenging high-quality data sets currently available for testing. These issues lead to inconsistencies among the experimental results in different papers in literature. In order to rectify those challenges, first, scene for the experiment is selected to be invariant to the parameters such as lightings. The scene should resemble the exact walkway with wide area so that this system has adequate information to detect those activities classified in it. Then, few videos are collected where persons are acted those selected activities of interest. The video collected is limited at the rate of 5 frames per second (fps) due to less memory. Other researchers suggested video with 1fps has the advantage in effectively human movement due to errant movement of human locomotion. The developed framework successfully detected human in all the videos collected at each frame thereby correctly recognised the suspicious human movement. Results are presented respective headings.

6.1 Pre-Processing

At pre-processing, human motion is detected using GMM and the detection is at each frame of the video. Figure 5 shows the human movement from the recorded video where the person walking in the area. Three frames shows in the Figure 5 collected at three seconds. Showing sequential pixels instead of seconds is inviable in this dissertation because distinguish movement only can be seen in a long period of time. Each frames in the Figure 5, showing the movement of human which is walking for three seconds. This makes each frame is collected at fifth frame of each second. Figure 6 shows the results from GMM where the ‘white’ pixels indicate the foreground and ‘black’ pixels representing the background. The foreground is the human movement which is interested in this thesis. Hence, GMM successfully detected human movement.

As GMM is sensitive to illuminations, this makes ‘white’ dots in the Figure 6 apart from the human shape. Morphology operations successfully eliminate those noises as in Figure 7. Figure 7 (a) shows the foreground with noises whereas (b) shows the result from morphological operation. Determining the suitable structuring element is done through systematic trial. After several attempts square shape of structuring element is finalised. Closing morphological operation is performed to the image using square structuring element in order to eliminate those white dots.

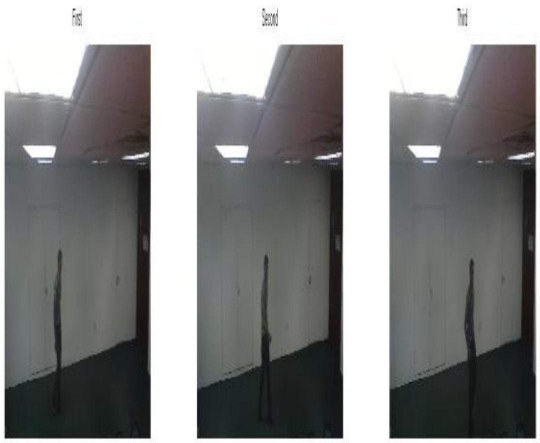


Figure 5: Human-movement

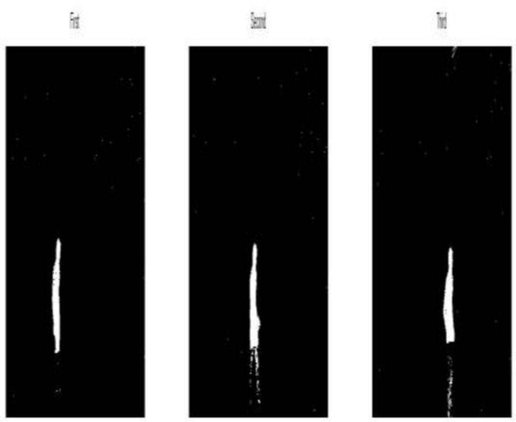
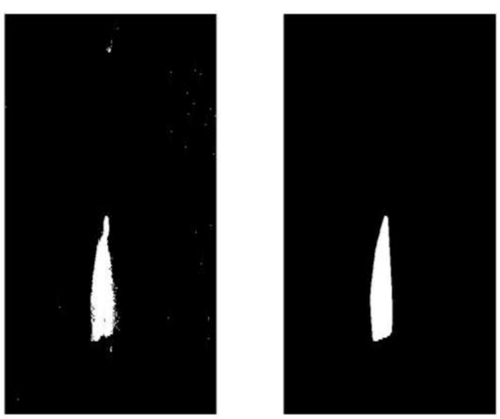


Figure 6: Results from GMM in grayscale



a) b)

Figure 7: Morphological Operation

6.2 Feature Extraction

In feature extracting, the 'white' pixels representing human is enclosed in a green bounding box as illustrated in Figure 8. The centroid of the bounding box is calculated by obtaining the size of the box. Eventually, a yellow dot representing the centroid is shown in the Figure 8. Moreover, there are two texts presented on the video player denoted as 1 and 2. Text labelled as 1 show the presence of human in the context. Second text shows the direction of human moving. As for the Figure 8, the person is moving towards the entrance; hence the text is showing 'PERSON IS MOVING TO THE ENTRANCE'.

Accuracy test is performed to determine the effectiveness of feature extraction in pre-process of this framework. The system yields high accuracy in detecting human using GMM. In order to test the accuracy of this system, different person are acting and being recorded in several videos with different types of movement, i.e. slow walking and fast walking. The system is able to recognise human in the test videos successfully. However, if the person is not moving in the context or the person moving not drastically, the system showing some issue in tracking them. Figure 9 shows the glitch in the system where green colour bounding box in the first frame is not moving as the person is move slightly. This shows the system is limited to certain level of movement. If the person performs minute movement, the framework cannot detect the movement. However, this problem will not affect the detection of suspicious movement. The reason is activity of interest did not consist of small movements.

6.3 Human Movement Detection

Before developing the algorithm for recognising suspicious movement, the scene for the scene is prepared as shown in Figure 10. The big rectangle with yellow colour shows the area for the person to walk. The small rectangle within the big rectangle is the ROI for the loitering. The ROI is set-up near the entrance (the door).

A) Loitering

Figure 11 shows the result of the scene of loitering. The first frame shows the person in the ROI and the timer is activated once the person is detected inside it. When the timer reaches 30 seconds, the system identified the person is loitering as shown in the second frame in Figure 11 by changing the bounding box to red colour. There is a text in red colour saying “PERSON IS LOITERING NEAR THE ENTRANCE” to signal the guard what type of activity is detected.

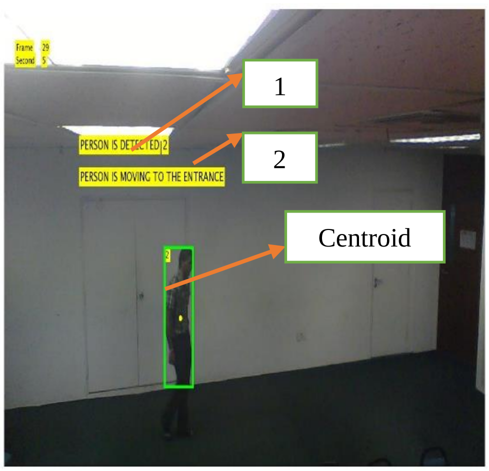


Figure 8: Bounding box enclosed human movement

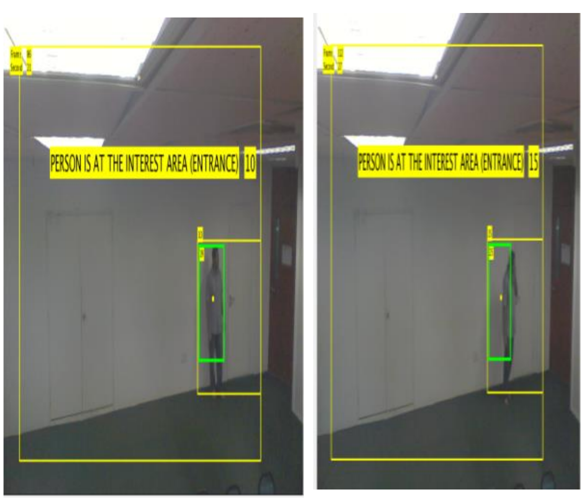


Figure 9: Person is performing minute movement

B) Looking Around

For looking around, person must be hanging around the place for more than 60 seconds, and the result is showed in Figure 12. Red colour bounding box surround the person indicating the person is looking around. The alert is aided with a text saying “PERSON IS LOOKING AROUND” in red colour.

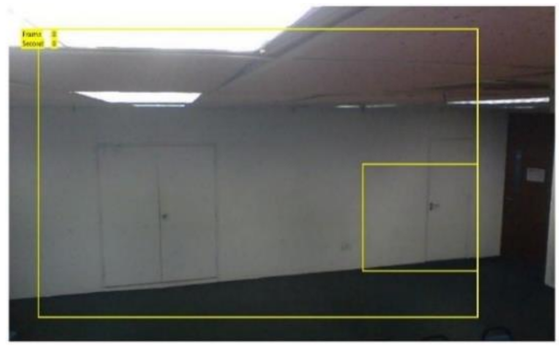


Figure 10: Scene for the experiment

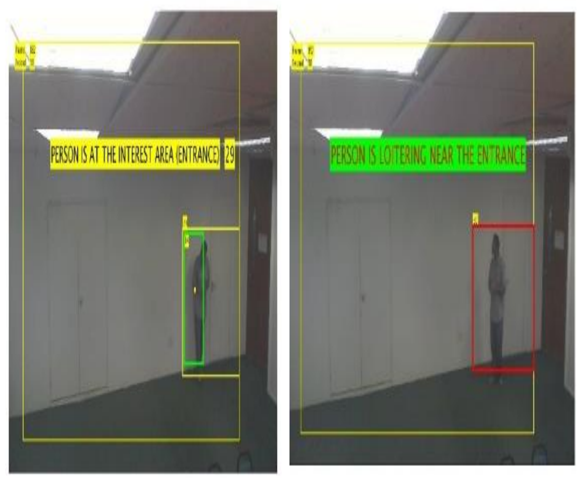


Figure 11: Person is loitering

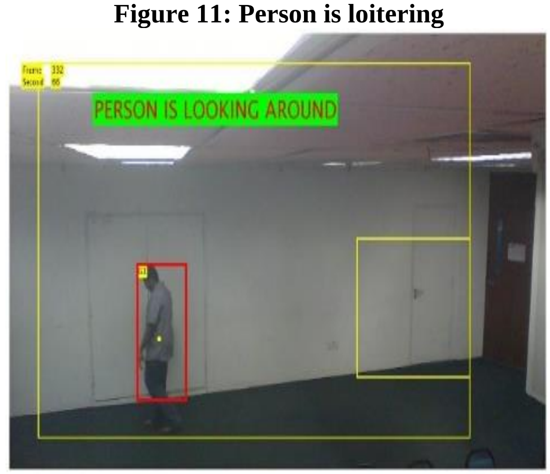


Figure 12: Person is looking around

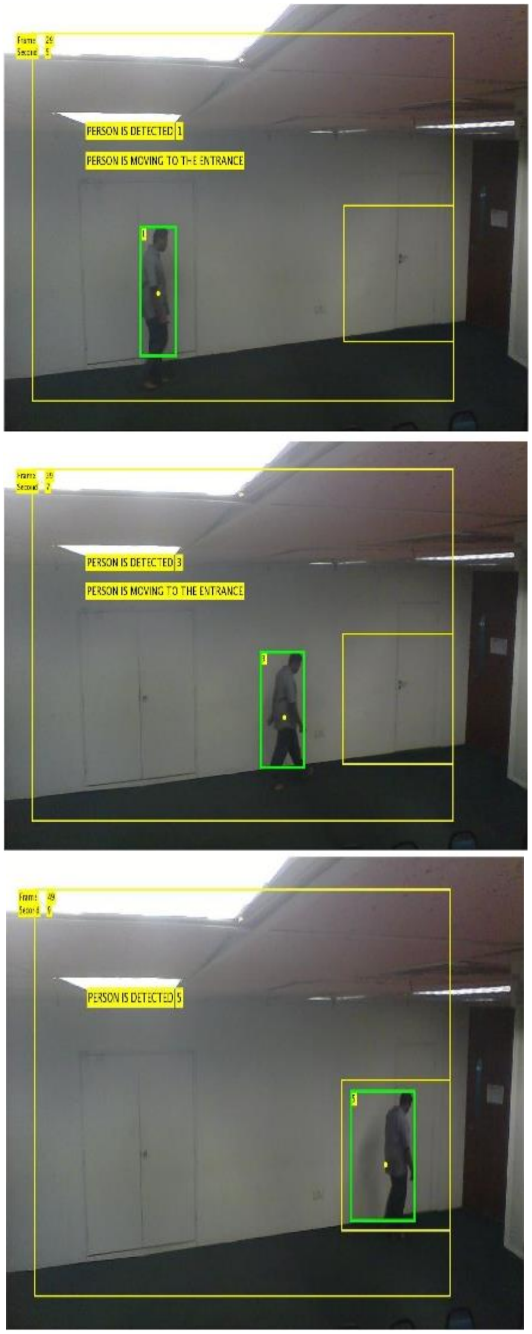


Figure 13: Normal human movement

C) Normal Human Movement

To distinguish activities of interest with normal activity, normal human movement is recorded to be tested. Effectiveness of a system also depends on accuracy of the recognising human-movement. Wrongly recognised movement will gives a poor framework. Hence, experiment is done to test the system whether can recognise normal human movement without giving any alert. Figure 13 shows the result where the person is walking towards the door as normal behaviour in the context. Consequently, as predicted any alerts is not given as indication of normal behaviour of human.

7 DISCUSSION

In the following, Table 4 shows the experimental results which are tabulated to show the qualitative and quantitative results respective to the behaviour of interest. From the result, the accuracy of the human detection using GMM is deviates for different activities of human. For normal walking, GMM successfully detected human in every frames. For looking around, the GMM also produces high accuracy in detecting human movement. However, for suspicious human movement like loitering, the GMM shows some glitch by giving low accuracy in detecting human movement. For event like loitering the person performs minute movement which makes the GMM difficult to identify the movement as foreground image. However, previous location of the person is stored in the system. This makes the system to react to the minute movement by acknowledging that movement as foreground movement.

Thus, accuracy did not affect the grammar-based approach in determining the movement of human is suspicious or not. The system still manages to make decision in determination of suspicious activity despite of low accuracy. This framework detects suspicious movement within a second. To be more precise, before the video goes to next frame, the system recognises a person’s activity and projects it in the next frame. Since the video is at the speed of 5fps, before the video reaches the next second, consecutive frame will project the results once the system recognised it. It is evident that this framework yields high speed.

Table 4: Experimental result

Scene	Behaviour	Accuracy	Evaluation
Person walking slowly towards the door	Normal	100%	Successfully detected
Person walking fast towards the door	Normal	100%	Successfully detected
Person is loitering more than 30s	Loitering only	53.5%	Successfully detected
Person is loitering more than 60s	Loitering and looking around	68.4%	Successfully detected loitering not looking around
Person is hanging around more than 60s	Looking around	95.1%	Successfully detected
Person is hanging around for more than 60s and overlapping the ROI	Looking around	100	Successfully detected

8 CONCLUSION

In this paper, grammar-based suspicious human movement recognition approach that depends on human detection and tracking is extensively investigated and the results have been presented. This paper begins with modelling the video in order to extract features about human movement. GMM used to detect and track human movement in the area. The result from GMM is represented in grayscale where 'white' pixel represents the human in the context. Although the results are inconsistent due to illumination and errant movement, using morphological operation those unwanted pixels are eliminated.

By using the bounding box, each frame of the video is analysed and features such as location of human and direction of human movement is extracted. Extracted features are used to develop grammar based algorithms to recognise suspicious human movement. The approach ensures high precision and high accuracy in real-time performances. Experiment was carried out on multiple recorded video acted by different person to depicts loitering and hanging around. The experimental results demonstrated successful detection of the various activities of interest.

9 REFERENCES

[1] Chan, S.C., Zhang, S., Wu, J., Tan, H, Ni, J.Q., Hung, Y.S. (2013) 'On the Hardware/Software Design and Implementation of a High Definition Multiview Video Surveillance System'. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*. Vol. 3(2). pp.248-262

[2] Eng, H.L., Toh, K.A., Kam, A.H., Wang, J., Yau, W.J.(2003) 'An automatic drowning detection surveillance system for challenging outdoor pool environments'. *IEEE International Conference on Computer Vision*. Nice, IEEE. pp 1-8

[3] Guo, J.M., Hsia, C,H., Liu, Y.F., Shih, M.H., Chang, C.H, Wu, J.Y. (2013). 'Fast Background Subtraction Based on a Multilayer Codebook Model for Moving Object Detection' *IEEE Transactions on Circuits and Systems for Video Technology*. Vol. 23(10). pp.1809-1821

[4] Hsieh, J.W., Hsu, Y.T., Liao, H.M., Chen, C.C. (2008) 'Video-Based Human Movement Analysis and Its Application to Surveillance Systems' *IEEE Transactions on Multimedia*, Vol. 10(3). pp. 372-384

[5] Huffington Post (2013). "*Boston Bombings Prompt Mixed Support For Increased Surveillance*" Available: http://www.huffingtonpost.com/2013/04/24/boston-bombings-surveillance-poll_n_3149576.html [Accessed 29 Apr. 2014].

[6] Lee, D.S (2005) 'Effective Gaussian Mixture Learning for Video Background Subtraction' *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Vol. 27(5). pp 827-832.

[7] Lepon, J. and Popkin, R. (2007). *A Study of CCTV at Harvard*. 1st ed. [ebook] Available at: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&so

urce=web&cd=1&cad=rja&uact=8&ved=0CCkQFjAA&url=http%3A%2F%2Fwww.eecs.harvard.edu%2Fcs199r%2Ffp%2FJanaRachel.pdf&ei=oPJfU633H-aViAfr7oHIBg&usg=AFQjCNHgepwkeC7Eac9Z79QVV9HgRwJQSw&sig2=XnZ7_KsGYAAm0nyRl0W7lg&bvm=bv.65636070,d.aGc [Accessed 29 Apr. 2014].

[8] Li, H., Cheng, X., Zhang, S., Wu, Z. (2013) 'RECOGNIZING HUMAN ACTIONS BY BP-ADABOOST ALGORITHM UNDER A HIERARCHICAL RECOGNITION FRAMEWORK'. *IEEE International Conference on Digital Object Identifier*. Vancouver. IEEE. pp. 3407-3411

[9] Moeslund,T.B., Hilton, A., & Kruger,V (2006) 'A Survey of Advances in Vision-Based Human Motion Capture and Analysis'. *Computer Vision and Image Understanding Special Issue on Modeling People: Vision-based understanding of a person's shape, appearance, movement and behaviour*. Vol. 104(2-3). pp.90-126

[10] The Wall Street Journal, 2013 "*In Praise of Surveillance Cameras*" Available: http://online.wsj.com/news/articles/SB10001424127887323309604578434712417328162 [Accessed 29 Apr. 2014].

[11] Turaga, P., Chellappa1, R., Subrahmaniam, V.S., Udrea, O. (2008) "Machine Recognition of Human Activities: A Survey", *IEEE Transactions on Circuits and Systems for Video Technology*. Vol. 18(11). pp.1473-1488