FRONT YARD ROBBERY ACTION DETECTION:

CLIMB-OVER-GATE ACTION ANALYSIS

By

Lim Kar Heng

Supervised By

Prof. Maylor Leung Kar Hang

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONOURS)

Faculty of Information and Communication Technology

(Kampar Campus)

JAN 2021

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: **Front Yard Robbery Action Detection: Climb-Over-Gate Action Analysis**

Academic Session: Jan 2021

I LIM KAR HENG

(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in

Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

(Author's signature)

(Supervisor's signature)

Address:

15, Lorong Berkat 2,

Taman Berkat,

09400 Padang Serai, Kedah.

Leung Kar Hang

Supervisor's name

Date: 15 April 2021

Date: 16 Apr 2021

FRONT YARD ROBBERY ACTION DETECTION:

CLIMB-OVER-GATE ACTION ANALYSIS

By

Lim Kar Heng

Supervised By

Prof. Maylor Leung Kar Hang

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONOURS)

Faculty of Information and Communication Technology

(Kampar Campus)

JAN 2021

DECLARATION OF ORIGINALITY

I declare that this report entitled "**Front Yard Robbery Action Detection: Climb-Over-Gate Action Analysis**" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : __________________________

Name : Lim Kar Heng

Date : 15 April 2021

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to my supervisors, Prof. Maylor Leung Kar Hang who has given me this bright opportunity to engage in a computer vision project. It is my first step to establish a career in computer vision field. I would like to thank him for his supervision, guidance, patience and suggestions throughout this project.

Finally, I must say thanks to my parents and my family for their love, support, and continuous encouragement throughout the course.

ABSTRACT

Many families will install motion-detection sensors for security and surveillance systems. However, these systems have their own limitation. The alarm can be triggered by any moving objects such as a passing animal. A guard is required to constantly checking on the screen of what being recorded to know if a robbery is occurring for a normal surveillance system.

In this paper, a surveillance system using computer vision and pattern recognition is proposed. The system will begin monitoring the front yard after it has been activated by the user. During the monitoring processes, the system will keep track of the action begin done and an alarm will be triggered when a climbing/jumping over gate or fence action is detected.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CCTV	Closed-Circuit Television
CNN	Convolutional Neural Network
GMM	Gaussian Mixture Model
k-NN	k-Nearest Neighbor
MSVM	Multiclass Support Vector Machine
OpenCV	Open Source Computer Vision
RLU	Rectified Linear Unit
YOLO	You Only Look Once

Chapter 1: Introduction

1.1 Problem Statement and Motivation

According to Numbeo, the world's largest crowd-sourced global database, Malaysia is the highest crime rate country in South East Asia, having crime index of 58.55, and this place Malaysia in rank 26 across 133 countries worldwide in 2020 mid-year. Crime index is an estimation of overall level of crime in a country, range from 0 to 100. Crime levels between 40 and 60 are considered moderate, but this means that crimes still happen quite frequently. According to the Department of Statistic Malaysia (2019), 60.8% of the violent crimes that happening in Malaysia were robbery for the year 2019. In order to prevent household robbery, many families will install motion-detection sensors for security and surveillance systems.

However, these systems have their own limitation. A motion-detection sensor is used to detect movement in those systems and an alarm will be triggered when a movement is detected. This is considered bad because the alarm can be triggered by any moving objects such as a passing animal. As for a normal surveillance system without motion sensors attached to them, they usually require a guard to constantly checking on the screen of what being recorded to know if a robbery is occurring, and then the alarm will be trigger by the guards. A robbery might happen while the guards is not paying attention on the screen, therefore the robbery cannot be prevented.

In order to overcome the limitation of current motion-detection sensors for security and surveillance systems, an AI-powered surveillance system can solve some of the problems the previous system has because it will reduce the rate of false alarm trigger as this system will be able to detect certain action from a robbery scene using computer vision and pattern recognition. The cost to hire a guard can also be omitted as this system does not require constant human monitoring.

1.2 Project Scope

To solve the problems stated above, the final product of this project will be a surveillance system that can detect front yard robbery action. The system will begin monitoring the front yard after it has been activated by the user. This surveillance system will be able to track multiple objects that is within its range. During the tracking process, the system will monitor the optical flow generated by those objects, and once a flow similar to the jumping action is detected, the system will then determine whether the flow is generated by human, if it is human, an alarm will be triggered.

1.3 Project Objectives

The main objective of this project is to develop a real-time surveillance system which will be better than the motion-detect sensor security or alarm system as this can reduce the false alarm triggered by any moving object. This real-time surveillance system also aims to reduce the delay in triggering an alarm. This is because the amount of human interaction required to trigger an alarm is reduced as it will automatically trigger alarm when a climb-over-gate action is detected.

In a nutshell, the objectives are as follow:

i. To develop a real-time surveillance system

- The system will be able to react to the activities immediately. Optical flow is generated immediately based on the video captured by the surveillance camera.

ii. To reduce the false alarm triggered

- This system will only trigger an alarm when a robbery-related action is detected. Those action and movement can be detected by using the optical flow calculated from each object tracked by the system.

iii. To reduce the delay in triggering an alarm

- This will eliminate as much human interaction required as possible in triggering an alarm.

1.4 Impact, Significance and Contribution

This surveillance system developed will be able to reduce the crime rate at the resident area since it will be difficult for robbery to happen when the household is being monitored by the surveillance system. It will indirectly improve the securities of the resident area and the people will feel safer as the crime rate reduced.

Next, this surveillance system can be considered as the improvement compared to motion-detect sensor security or alarm system as the alarm will only be triggered when certain action is detected unlike where in motion-detect system where alarm is triggered every time a motion is detected, thus reducing the rate of false alarm triggered.

Besides that, this real-time surveillance system requires little human inputs, and it does not require a person to constant monitoring what is being captured by the surveillance system. This is because as mentioned before, the system will be able to trigger an alarm automatically when the climb-over-gate action is detected.

1.5 Background Information

With the technology keep advancing every day, technologies such as image processing, computer vision and pattern recognition emerged. These technologies can be integrated into the current security system to make them more effective. According to Eduardo and Gelson (2005, p. 891), digital image processing is the manipulation of image using computers, and can be categorized into different processes, such as image enhancement, image restoration, image analysis, and image compression. Image analysis is the focus of this project. It allows information to be extracted automatically from an image while it is being processed by segmentation, edge extraction or texture and motion analysis. With computer vision, it allows computers to see, observe and understand based on the image processing techniques.

Optical flow is the pattern of a motion of an objects, surfaces, and edges in a scene that is caused by the relative motion observed in the scene. It is mainly used in motion estimation. The optical flows are generated by calculating the motion between two frames that are taken continuously.

Chapter 2: Literature Review

Robbery are usually being done by criminals and most robbery cases have similar actions or behaviours that can be observed such as people pointing guns at others, wearing black mask, and the victims raising their hand while crouching. Abnormal behaviour recognition techniques and deep neural network is used to classify these actions and determine whether a robbery is happening.

Beside classification, object tracking is also needed for surveillance system as it allows the system to track each object that appear in the scene. It is very useful to track the intruders and triggering an alarm when they are entering the house by climbing over the gate.

This chapter includes the materials, review of existing video surveillance system, abnormal behaviour recognition techniques, object tracking techniques, deep neural network, and optical flow techniques.

2.1 Abnormal behaviour recognition for intelligent video surveillance systems: A review by Mabrouk, Zagrouba (2017)

Mabrouk and Zagrouba (2017) had done a review on abnormal behaviour recognition for intelligent video surveillance system. According to them, human action recognition has one problem, and it is abnormal behaviour detection. The managing process is harder due to the increasing number of surveillance cameras, it becomes very difficult to supervise all monitors, and the abnormal events are infrequent. Therefore, an intelligent video surveillance system that can automatically detect an irregular behaviour is in huge demand right now.

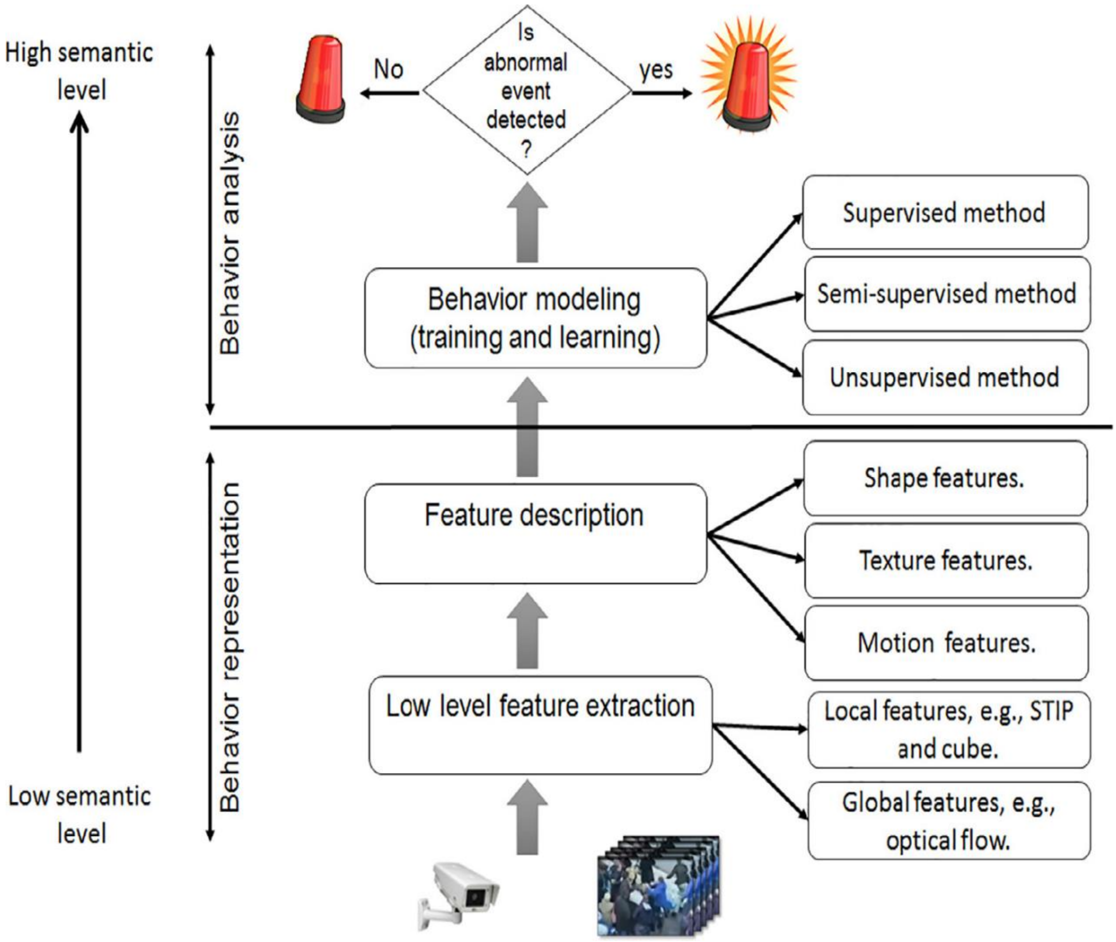


Figure 2.1.1 Intelligent video surveillance system (Mabrouk, Zagrouba 2017)

The main objective of an intelligent video surveillance system is the identification of an unusual occasion from large number of videos efficiently to avoid accident from happen. This job requires two processing level as shown in figure 2.1. First, it is the low-level features which is used to identify the desired area in the scene and extract it. Then, simple description is generated on the low-level features. The second level of this task is to provide the data about the human behaviour and decides whether it is normal or not.

The low-level processing step of the behaviour analysis is the behaviour representation. The objective of behaviour analysis is to capture relevant features that can be used to describe the target object in the recording. In this process, an unusual region in the scene is detected. Then, this region is given an explanation. This level of processing is the most challenging and difficult as it will significantly influences the understanding of the object behaviour. To find a suitable feature that are unaffected by transformation such as the changes in the surrounding or on the object itself is the major challenge in behaviour representation.

The low-level processing stages only permit identifying and describing the object in the video, it cannot determine if the behaviour is normal or not, which is why the second level of processing is required. Recognizing an irregular behaviour depend on the classification model or method used, and those methods can be divided into 3 different categories, supervised, semi-supervised and unsupervised.

Supervised methods purpose is to model a normal and an irregular behaviour through labelled data. They can detect specific irregular behaviours learned during the training phase of the models. For semi-supervised methods, they only need regular video for training and can be separated into 2 approaches, rule-based and model-based. The first approach will deliver a rule based on regular behaviour, any sample that does not follow the rule is treated as irregular. The second approach, model-based method, the model representing a regular behaviour is deviated into an irregular behaviour related to an instance. For unsupervised methods, the model is trained with both regular and irregular behaviours using statistical information obtained from unlabelled data.

2.2 Vision-Based Human Tracking and Activity Recognition by Bodor et al. (2003)

Bodor et al. (2003) had developed an automated surveillance system to track pedestrians and detect any abnormalities such as when a pedestrian enters or being near a “secure area”, moves above average walking speed, stay in the area for some time or falls down. This system is developed in two part, which is pedestrian images capturing and tracking and pedestrian activity recognition based on velocity and positioning.

The pedestrian images capturing and tracking part of the system is built on a framework developed by Harini Veeraraghavam. With the use of Gaussians for segmentation and Kalman filter for tracking, this system can track objects appearance in a video sequence. This step aims to segment and extract each pedestrian’s image from all appearances in the video sequence, and then track the pedestrian and an image snapshots of the pedestrian’s motion will be generated as shown in the Figure 2.2.1.

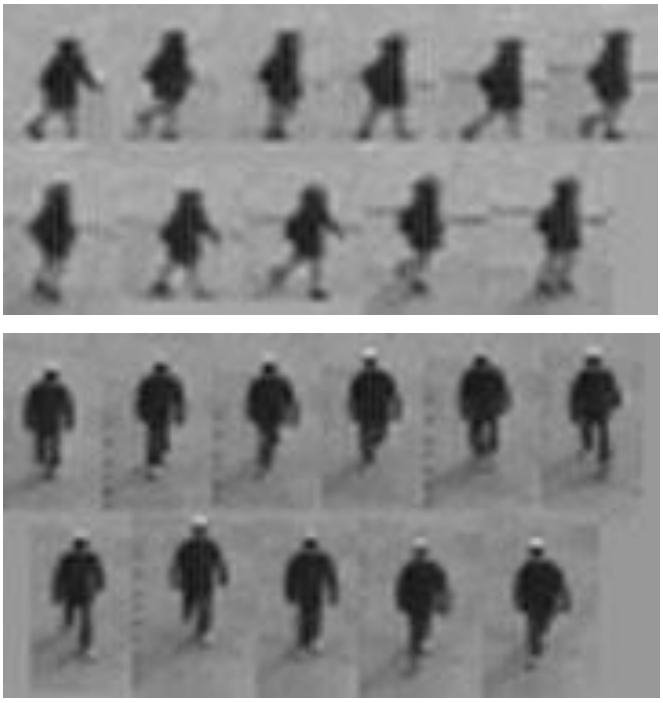


Figure 2.2.1 Sample of snapshot of pedestrian’s motion (Bodor et al. 2003)

For the second part of the surveillance system which is pedestrian activity recognition based on velocity and positioning, it will estimate the pedestrian’s activities based on the speed and its position. According to Bodor et al., measuring these values has several benefits compared to normal motion analysis. They said that these measurements are obtained in real time and do not affected by noise or poor image quality. Furthermore, the type of motion is not important as one of the abnormalities this system want to detect is when a pedestrian enter an off-limit area. A warning should be trigger whenever someone is inside that area, no matter what action is performed by that person in that area. The off-limit area is decided with the help of human input.

Figure 2.2.2 Off limit area on the top right (Bodor et al. 2003)

This system has its problems and limitations. Sudden changing in the lighting of the outdoor scene caused by the sunlight being blocked by cloud will causes some error in the system. 15 seconds is required for the background segmentation method that is used by the system to adapt to the changes in lighting, and this will cause blobs to appear even though there is no person or object at the scene. Another limitation of the system is that it cannot differentiates between objects which are moving at the same speed through different ways, such as cycling or running.

2.3 An Intelligent System to Detect Human Suspicious Activity Using Deep Neural Networks by Ramachandran et al. (2019)

Ramachandran et al. (2019) has proposed a framework to detect suspicious human action and track those that are doing such suspicious action. This proposed model is trained with a huge set of labelled data which consist of different categories from the Convolutional Neural Network (CNN) with multiple layers such as convolutional layers, activation layer and pooling layer. Figure 2.3.1 show the flow of the proposed system.

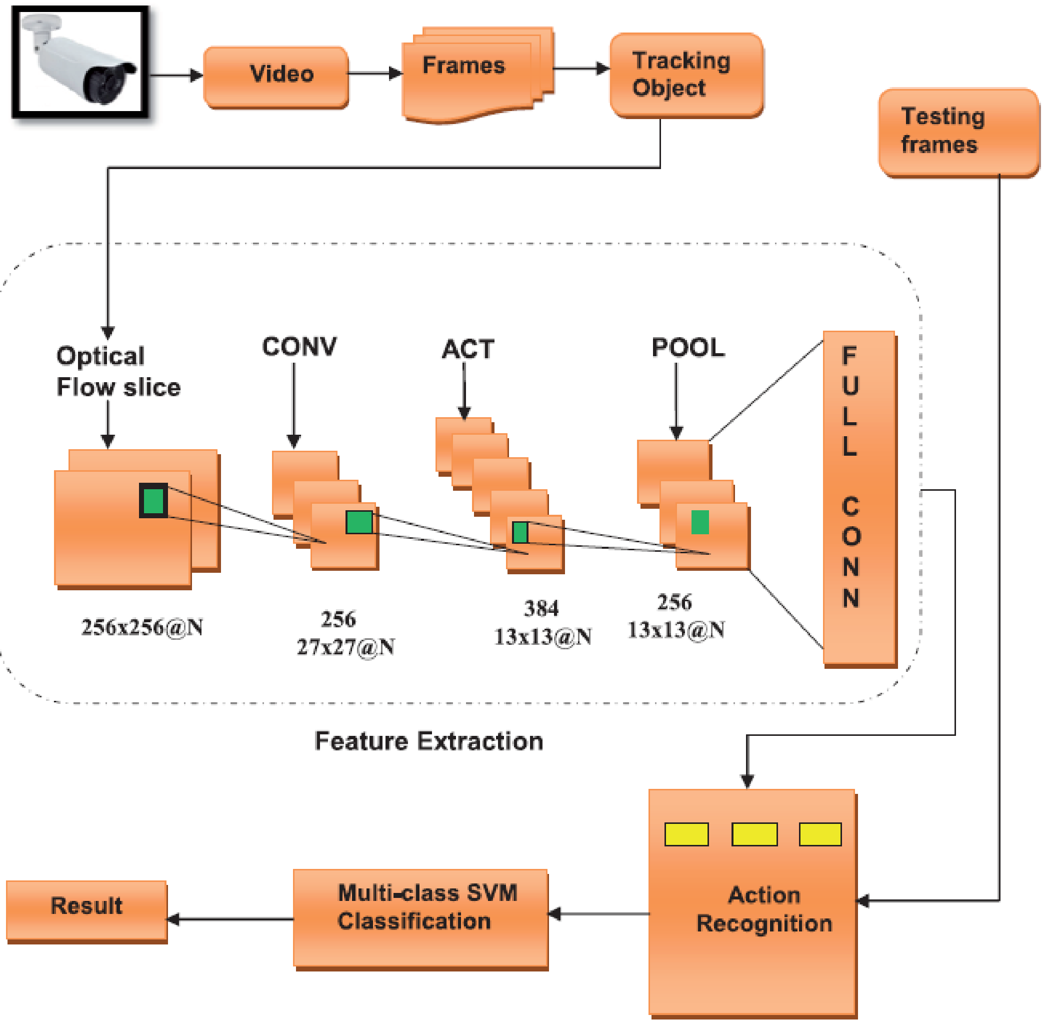


Figure 2.3.1 Suspicious Human Action Detection System (Ramachandran et al. 2019)

The system first starts with video inputs. This input is divided into frames and noises is removed from input video by converting RGB frame into Gray scale frame. Next, optical flow is used in tracking the object. Optical flow is a pattern of an object motion that is generated by comparing successive frames. It is a two-dimension field where each vector shows how a point.

After the tracking phase, it is time for feature extraction. Due to its simple structure, CNN is chosen as the model to be used. This process is done by passing the input into three different layers of CNN. The first one is a Conv Layer, which will perform a dot function between the input and its filters. Next is Relu layer, it will generate an activation value that are required by the CNN model based on the input provided by previous layer. After that, an image is produced as shown in Figure 2.3.2 using a max pooling layer, where a maximum value from each part of the image is summarize into a smaller image. The FC layer is used to predicts the output based on the summarization obtained on the pooling layer.

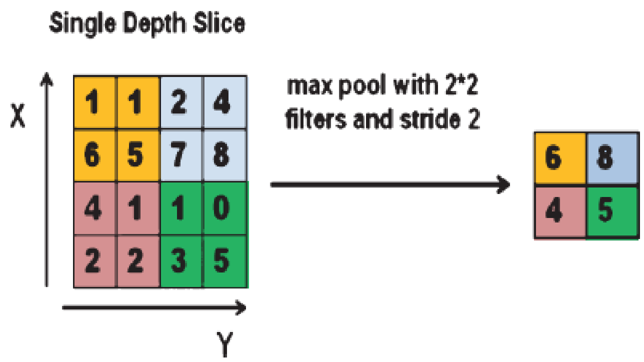


Figure 2.3.2 Pooling Layer (Ramachandran et al. 2019)

After the feature extraction of the images is completed, those predicted images is given to Multiclass Support Vector Machine (MSVM), to train it to categorizes the recognized actions. The advantage of this system is that it combines both CNN and MSVM for classification. The performance of the proposed method is shown in the Table 2.3.1.

Actions	Classifiers											
	KNN			Linear SVM			Random forest			Proposed CNN+MSVM		
	Accuracy (%)	Predicted speed (obs/sec.)	Trained Time (sec.)	Accuracy (%)	Predicted speed (obs/sec.)	Trained time (sec.)	Accuracy (%)	Predicted speed (obs/sec.)	Trained time (sec.)	Accuracy (%)	Predicted speed (obs/sec.)	Trained time (sec.)
Fighting	62	31	20.945	82	39	21.015	88	38	25.05	95	35	30.05
Boxing	60	40	28.275	78	32	23.245	84	30	27.20	90	39	24.14
Clapping	58	38	24.285	70	38	26.242	80	34	26.28	86	32	28.6
Running	54	30	21.032	72	29	21.495	84	28	28.44	90	28	29.94
Robbery	68	31	25.542	75	36	28.674	90	36	25.42	92	40	28.674

Table 2.3.1 Performance of Classifiers (Ramachandran et al. 2019)

Several datasets are used in the training process to determine the performance of each classifier. The database KTH contains both indoor and outdoor environment for human activities such as running, hand waving and clapping. BEHAVE database is also used and it contains several situations of people interactions. The data in this database is in avi format and each second has 25 frames. Another database that is used is FALL which consists of 30 falls and 40 activities of daily living in indoor environments, to a total of 70 sequences. This system also used Pickpocket database during the training process. The data from this database are all collected from the real-world events in the open areas such as bus station.

2.4 Automatic Video-Based Human Motion Analyzer for Consumer Surveillance System by Lao et al. (2009)

Lao et al. (2009) has proposed a framework for study of human action from a surveillance video captured by a user camera. The main focus of this paper is the automated analysis of human behaviour, with the objective to explore a more effective algorithms for usage at home. Figure 2.4.1 shows the block diagram for the surveillance video analysis system.

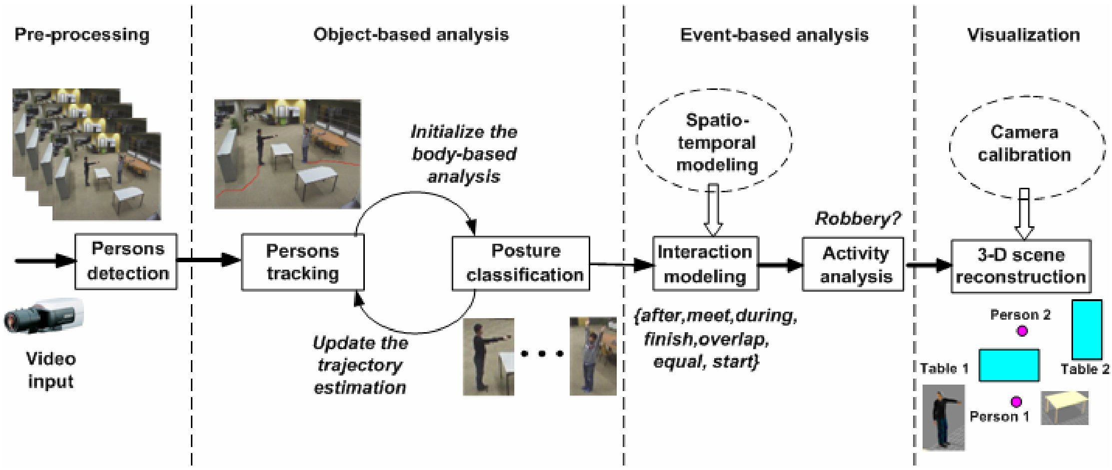


Figure 2.4.1 Block Diagram for Surveillance Video Analysis System (Lao et al. 2009)

The first step of this system is multi-person detection. A pixel-based background subtraction is performed, generating a probability density function for each pixel. A pixel from a new image will be considered as a background if that pixel's value is can be defined by its density function. The Gaussian Mixture Model (GMM) is used for background subtraction. After that, k-Nearest Neighbour (k-NN) is used to recognize people. This classifier uses area of the bounding box attached to each object to perform classification. The non-human objects and noises can be removed using this way.

Next step in this system is person tracking, called trajectory estimation. This step is performed along posture classification. Each distinct appearance model is represented by using a colour histogram that generated from mean-shift algorithm. For each new person entering the scene, a new histogram model is calculated. In the following frames to track that person, the person object is shifted to the location whose histogram is the closest to the previous frame.

The understanding of interaction between people involved in the scene is important in a multi-person event in order for the event analysis to work. The casual and time-overlapping relations of the two persons' gesture changes will help in defining the temporal constraint between two person interactions. To represent temporal relationships better, temporal logic is applied. Seven temporal relationships are showed in the set TR, which consist of *{after, meets, during, finishes, overlaps, equal, starts}*. These can be linked with different sub-events after each distinct event is analysed. For example, the posture "pointing" can be infer to a robbery event. Other than "pointing', "raising both hands" and "squatting" can also be used to infer that a robbery scene is happening.

For the last part of this system, in the visualization step, a 3D scene will be reconstructed. After performing camera calibration either with automatic or manual, the 2D-3D mapping can be implemented. The image captured can be described in a 3D domain. This reformation is useful for a better understanding on the scene. The camera calibration can provide a geometric information and it allows the mapping of the points in image domain to the real-word coordinates. Figure 2.4.2 shows an example of a scene reconstruction. In this system, the human height is not a required information since the human behaviour is just based on the person's course on the ground.

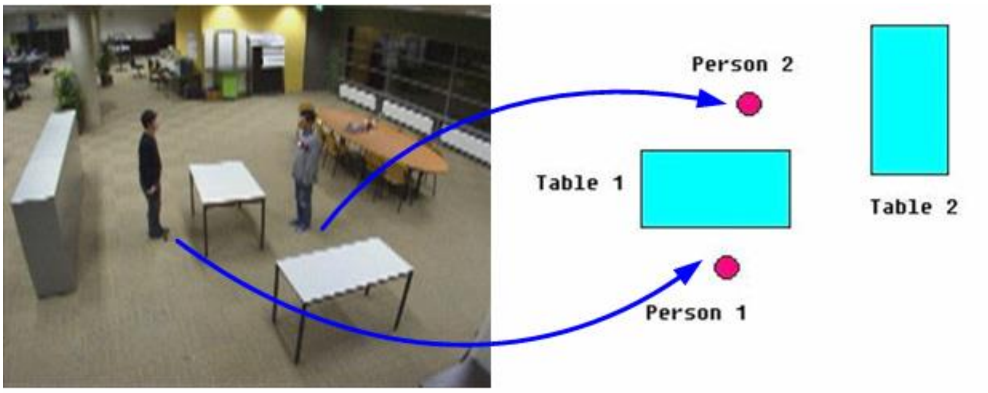


Figure 2.4.2 Example of Scene Reconstruction (Lao et al. 2009)

The limitation of this system is that occlusion is not yet currently. However, to solve this problem, several cameras can be set up to capture the same scene from different viewpoints.

2.5 Optical Flow Modeling and Computation: A Survey by Fortun, Bouthemy and Kervrann (2015)

Fortun, Bouthemy and Lervrann (2015) did a survey on optical flow modelling. They proposed a classification of the main principles of the existing methods, and focus on the modelling aspects, practical matters and the limitations of each introduced method.

According to them, optical flow is the displacements of intensity patterns. Although optical flow is generated by the relative motion that is observed on the object from the observer on the scene, it only represents the intensities of the motion in a 2D plane, but not represent the actual 3D motion in the scene. The intensity of moving pixels will remain constant during the motion. Optical flow provides many fundamental information applicable in wide range of fields.

The two main visualization techniques are shown in the Figure 2.5.1.

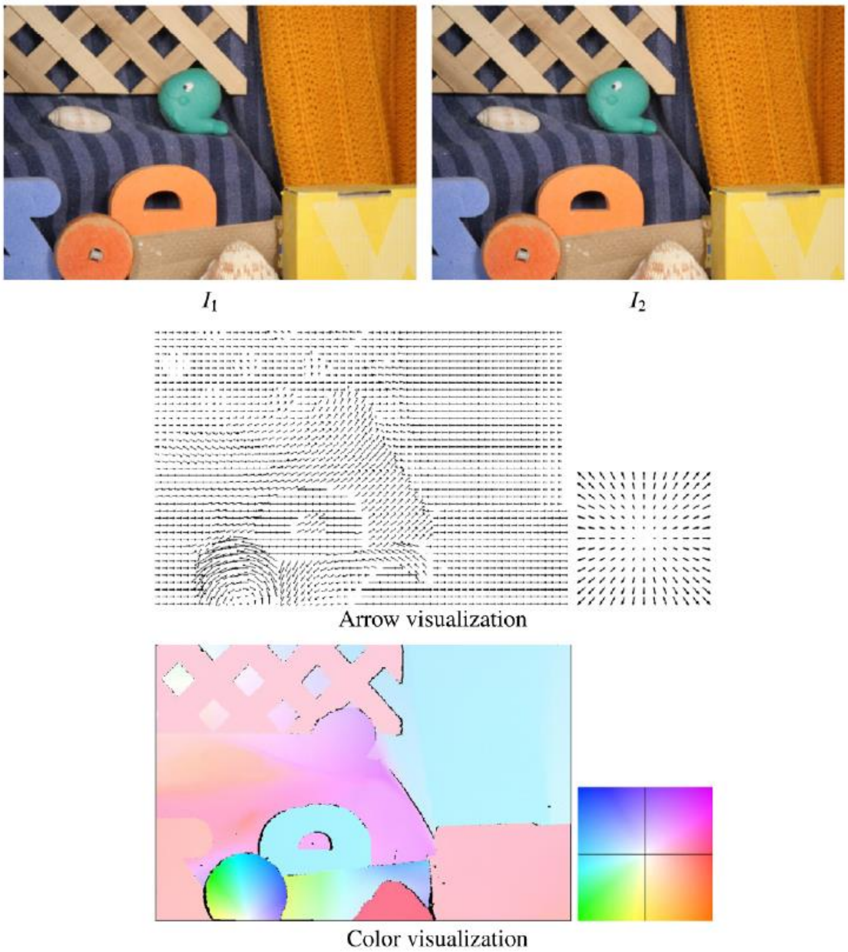


Figure 2.5.1 Two type of visualization (Fortun, Bouthemy and Kervrann 2015)

The arrow visualization represents directly on the motion vectors and it give a good insight of the motion. The colour code visualization applies a colour hue to track and saturation to the scale of the vector. The colour allows for a better visual observation of subtle differences between neighbour motion vectors.

The brightness constancy assumption is used for a scenario where imperfect photometric of an actual motion in the scene. A counterexample is where with the existence of moving light source in a static scene, the brightness changes produced can result in a generation of optical flow. This will generate noises in the information obtained. Several image transformation methods can be used to reduce the noises. The first one being image smoothing. Fortun, Bouthemy and Lervrann noticed that Gaussian smoothing is applied by most methods as a first step during pre-processing to reduce the influence of noises. Another way to work on the brightness changes is to make use of the texture components of the image. First the structure part of the image is obtained by reserving smoothing, and then obtain the texture part by subtracting the structure part with the original image.

The next problem in optical flow is occlusions. Occluded regions are a set of pixels that become unobservable due to the moving objects in the next frame. These pixels have no related points in the next frame which result in the motion is not visible at their locations. Occlusion detection has been used to solve this problem. It estimates the occlusion map along with the motion field by using an alternative optimization scheme. After that, occlusion filling is done by using the motion vectors that are related to the image inpainting. Image inpainting has two different category, diffusion-based and exemplar-based. Diffusion-based occlusion filling depends on the diffusion process to propagate the motion from non-occluded region to those that are occluded. It is not suitable to be used in an image with large missing region. Exemplar-based will attempt to design a valid data cost from the motion vector that matched the non-occluded pixel.

2.6 Design and Implementation of An Optical Flow-Based Autonomous Video Surveillance System by Fonseca et al. (2008)

Fonseca et al. (2008) implemented an autonomous video surveillance system based on only optical flow calculations. They discovered new ways for managing segmentation, depth estimation, tracking and object classification. Figure 2.6.1 shows the architecture of the surveillance system. Based on this architecture, all the information and data are shared at the system level. All information is obtained from optical flow calculations. The aim of this design is to reduce the number of algorithms implemented and increase the data sharing between modules.

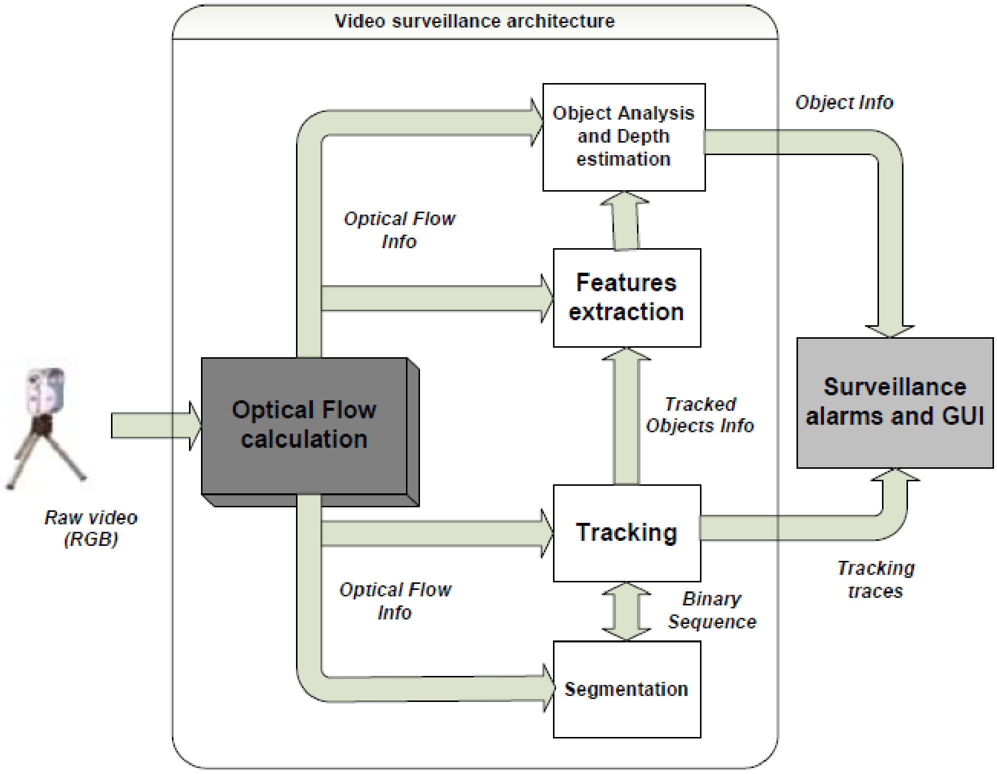


Figure 2.6.1 System Architecture (Fonseca et al. 2008)

Fonseca et al. decided to select Lucas-Kanade algorithm for this implementation as it can provide an acceptable result with the lowest processing time. For the segmentation module, it will use the result from optical flow calculations to determine which part in frame is part of the foreground and which is background. Figure 2.6.2 shows the average optical flow vector length per frame.

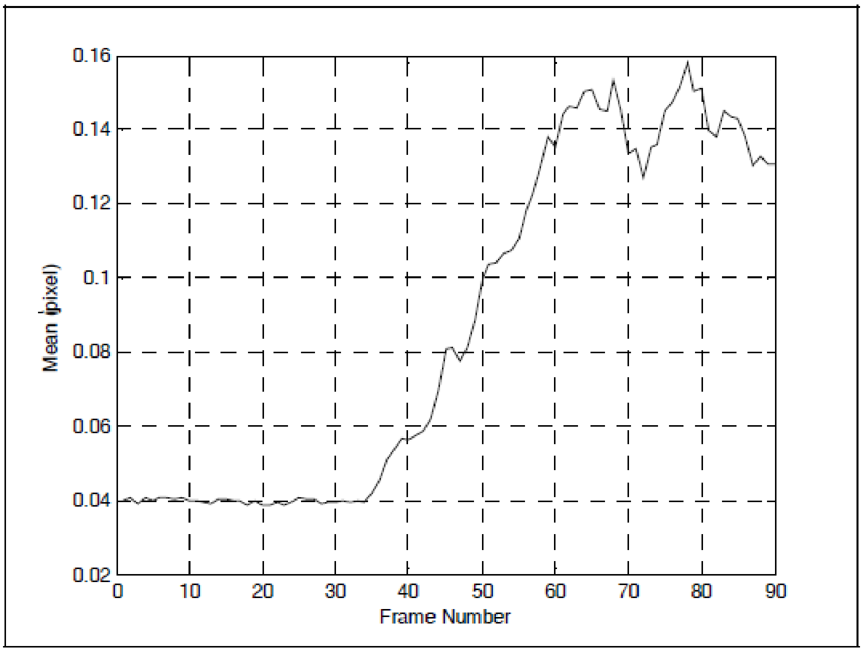


Figure 2.6.2 Average Optical Flow Vector Length per Frame (Fonseca et al. 2008)

The first 30 frames consist only the background, that is why the vector length is on constant 0.04. After frame 30, a moving object appear and this result in the optical flow vector length increases. These changes are used to separate background and foreground using a threshold value. That object is that tracked with its optical flow vector as a feature for tracking. The number of optical flow vectors per object will reduce when the object is further away from the camera, using this in along with the mean x and y optical flow values per object, an estimation in the 3D trajectory of an object can be obtained.

The most noticeable limitation when using optical flow is when dealing with non-moving objects. These static objects will not generate any optical flow. However, this system can detect when an object is stop moving and no longer be tracked.

Chapter 3: System Design

3.1 Design Specification

3.1.1 Methodologies and General Work Procedures

The proposed system will identify the foreground object and detect whether a human from outside the house is entering the front yard through climbing the gate or the fences. This system needs to detect all moving objects in the scene. Next, the object detected will be tracked to monitor their movement. Once the object seems like it is making a jump over or a climb over action over the gate and fences into front yard, YOLO will be started to identify the object. If the object making that action is identified as a person, an alert will be triggered. OpenCV and YOLO will be used to develop the system. The programming language used will be C++ as it is the language OpenCV library is written in.

3.1.2 Tools to use

1. C++

C++ is a programming language created by Bjarne Stroustrup as an extension of the C programming language. It is selected as the language to use for developing the system because it has access to the OpenCV library.

2. Open-Source Computer Vision (OpenCV)

Library for C++ language which main function is aimed at real-time computer vision. It is used for segmentation, motion tracking, object identification and many more. OpenCV also contain optical flow function that can be used.

3. You only look once (YOLO)

YOLO is a real-time object detection and recognition algorithm. It is faster and more accurate compared to another object detection algorithm. This algorithm applies a single neural network to the whole image and predicts the bounding boxes and the object detected within that boxes. Darknet is a framework used to implement YOLO.

4. Visual Studio

Integrated Development Environment (IDE) for programming in C++ language.

3.2 System Flow

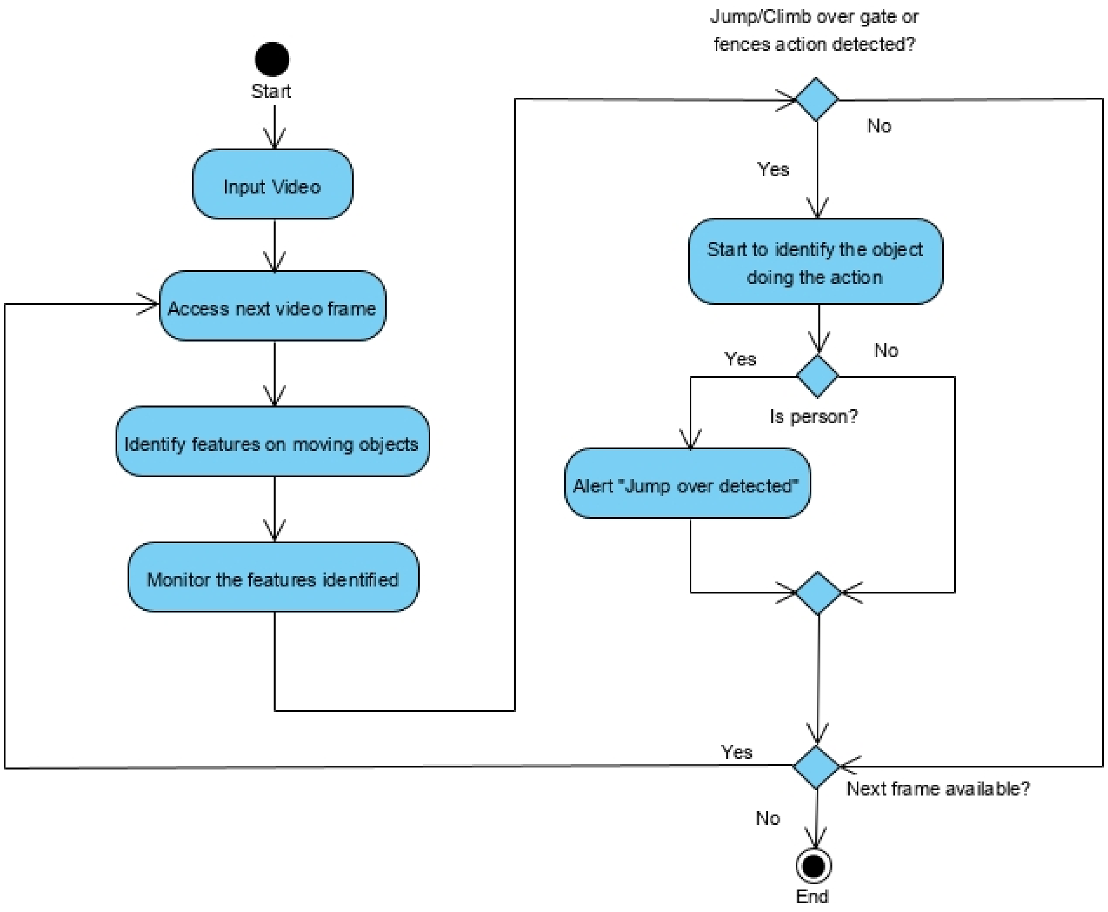


Figure 3.2.1 System Flow

The figure above shows the flow for the surveillance system that will be used to detect jump-over/climb-over gate or fences action. Firstly, a video is provided as input for the system, and during every 10 frames of the video, the features on the moving object is identified. Next, the system will monitor the features identified and calculate optical flow for each of them. Based on the flow generated, the system will decide whether a jump-over/climb-over gate or fences action is occurring. Once the specified action is detected, darknet will be triggered to use YOLO to identify the object in the frame. If a person is detected for doing that action, an alert will appear. All these processes will be repeated until it reaches the last frame of the video.

3.3 Assumptions

Several assumptions are made to make sure the system perform properly:

a. The camera must be placed in the corner of the front yard so that the top-down view of the gate and fences can be capture perfectly.
b. The action must be quick, if the person is climbing really slow no optical flow will be generated.
c. Each household will require different parameter setup for the system.
d. The person must do a direct dropdown from gate/fence and not leaping forward.

3.4 System Pseudocode

<u>System Setup Stage</u>

Step 1: The input video is viewed before hand to know around exactly where the location of the gate/fences is. Then these locations are manually specified in the program before it start up.

Step 2: Some parameters are initialized. Those parameters include the confidence threshold for object detection, and threshold for triggering the alert.

<u>Pre-Processing Stage</u>

Step 1: The input video is read and resized to a smaller scaler so that it is faster to be processed.

Step 2: The YOLO model configuration and weights is loaded into the Darknet framework.

Step 3: Some random colors is generated for drawing the optical flow later on.

<u>Processing Stage</u>

Step 1: The frame is read from the input video one by one, and for the first two frame, an absolute different between them is calculated to act as a mask for the feature point detection function.

Step 2: For every ten frame or when the amount of feature point is low, the feature point detection function will be called to update the points.

Step 3: Optical flow is calculated for those points generated by the function.

Step 4: If multiple vertical movement is generated from optical flow calculation, YOLO is started to identify the object generating those movement.

Step 5: Once YOLO identify a person within the frame the vertical movement is detected, it will check if those action is within the boundary box of the human.

Step 6: When most of the points is within the boundary box, an alert saying "JUMP OVER DETECTED" will appear.

Step 7: If nothing is detected, the program repeats from step 1 of the processing phase again until no new frame is available.

<u>Post-Processing Phase</u>

Step 1: All the display video and input are released, and the system shut down.

3.5 Timeline

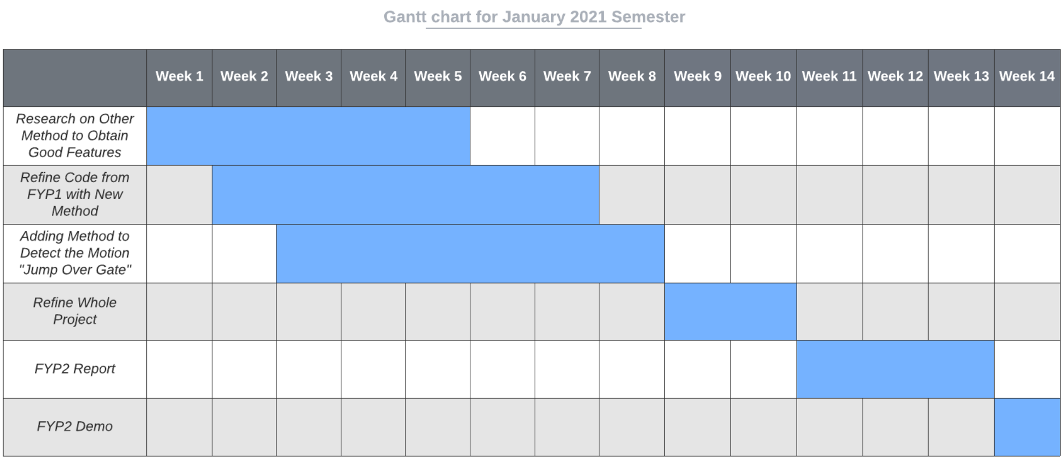


Figure 3.5.1 Gantt Chart for January 2021 Semester

Chapter 4: System Implementation

4.1 Current System

4.1.1 System Initialization

Several initializations are required to be inputted by user manually by looking at the scene captured. Those input include:

i. Above where on the scene the gate and fence located.
ii. Number of points of vertical movement from optical flow to be consider a jumping action

Figure 4.1.1.1 Scene 1

In figure 4.1.1.1, the gate and fence are located above the blue line (the line is drawn for explanation, it is not in the actual system). The user is required to manually find the find the location and enter it into the system.

Figure 4.1.1.2 Scene 2

In figure 4.1.1.2, the gate and fence cover the whole scene, the user just needed to input the max height as the location for gate and fence. Since the rest of the implementation is mostly the same, most of the time Scene 1 will be used in explanation.

4.1.2 Optical Flow and Point Detection

In order to obtain the feature points only for moving objects, two continuous frames are used in a subtraction and modulus is done on it so only the positive value is obtained. If those frames have moving object in it, a blob will be generated as shown in figure 4.1.2.1.

Figure 4.1.2.1 Blob Generated by Subtracting Two Continuous Frames

This blob will be passed to the goodFeaturesToTrack function as the mask. Only those features on the mask will have a point generated to be tracked. This process will be done every 10 frames or when the total feature points is low so that the features to be tracked is updated constantly.

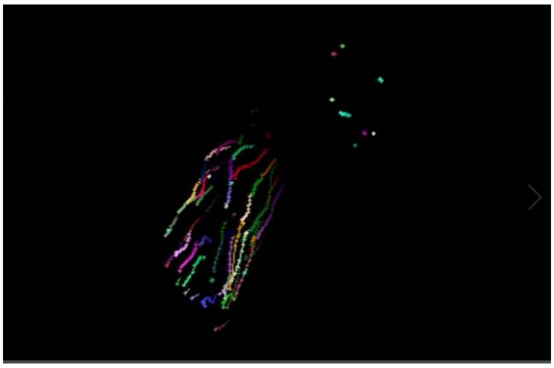


Figure 4.1.2.2 Optical Flow

These points will be passed to calcOpticalFlowPyrLK for optical flow calculations. Optical flow will be calculated every frame.

4.1.3 Action Detection and False Alarm

Figure 4.1.3.1 Jump Over Detected

Every time the optical flow calculations are completed, the system will check for any vertical movement generated. This can be done by using the y-axis value calculated in the optical flow. If the y-axis value for the new point minus old point is greater than 0, it means that a downward action is happening. When multiple downward action is detected, YOLO will be activated.

YOLO will start to detect if there is any person appearing in the frame. If a person is detected, the system will check if the vertical flow is within the boundary box detected. If it was, it means that a person is jumping over, and an alert will be triggered.

In Section 4.1.1, it is mentioned that the location of the gate and fence is needed to be manually inputted. If the user did not do this, false alarm will occur as seen in the figure below.

Figure 4.1.3.2 False Alarm

Multiple vertical flow is detected within the frame, and without specifying where the gate/fence is located, the system will assume that this is a jumping action and treat it like one. This false alarm can be solved by telling the system where the gate and fence are located, as those places are usually where the jump over action will occur, so the system only perform checking on those location.

4.2 Previous Attempt

Before reaching the current system, dense flow has been tried to see if the performance will be better. Dense flow is where it tries to calculated the optical flow for each pixel in every frame unlike the one currently used where optical flow is only generated on the specified points.

Figure 4.2.1 Dense Flow Result

Since dense flow is computation heavy, so another method to detect person is used to replace YOLO as YOLO is also computation heavy. The method used to detect person is called HOGDescriptor. The accuracy of this method is not as good as YOLO, but it is a tradeoff for a better speed compared to YOLO.

In the end this attempt is discarded as the performance of this attempt is worst compared to the current system. It has a FPS of around 18, and the accuracy is not as great as the current system.

Chapter 5: Experimental Results and Discussion

5.1 Result on Scene 1

Figure 5.1.1 Optical Flow generated by the Person Walking into House

As shown in Figure 5.1.1, feature points are only detected on the moving object, on the walking person and the closing door. Optical flow is calculated for those points only. Notice that the flow generated by the walking person is not triggering any alert despite that the flow generated is similar to a jump over action because the system will only trigger an alert if the action detected is around the gate or fence.

Figure 5.1.2 Alert Triggered by the Jump Over Person

In Figure 5.1.2, a person can be seen doing the jump over gate action. Multiple vertical optical flow is detected by the system so YOLO is being activated to check if a person exists in that scene, and once a person is detected, it will check if the flow generated is within the bounding box of the person. If all condition is meet, an alert will appear.

5.2 Result on Scene 2

Figure 5.2.1 Optical Flow Generated by a Crowd

Figure 5.2.1 shows a moving crowd generating optical flow. The gate and fence cover until the whole scene so all the flow generated in the scene will be checked.

Figure 5.2.2 Alert Triggered

A person is jumping over the gate in figure 5.2.2. The same processes happen in the system, multiple vertical flow detected lead to activation of YOLO and then checking if the flow is within the boundary box of the person. If all conditions are meet, alert is triggered.

Chapter 6: Conclusion

6.1 Project Review

In this paper, Malaysia crime rate is studied, and it is found that robbery is one of the highest crimes that is happening in Malaysia. Therefore, a surveillance system capable of detecting intruders from outside of the house is needed to prevent this from happening. This paper proposed a surveillance system that can detect the jump/climb over gate and fence action.

The system developed can react to the activities immediately. The delay in triggering an alarm is also reduced because of that. The system also greatly reduced the false alarm triggered by normal motion-detection sensor because it only react to the specified action that has a high chance of indicating a robbery will happen. With these, all of the objectives stated in the beginning is fulfilled.

Several problems have surfaced during the development process. The angle of the camera is important in detecting the jump over action, so only a few of the sample video can be found to be used as a testing input. Next, the majority of the references found in the internet is outdated or it is written in another programming language, so more time is needed to perform study in order to develop the system.

6.2 Future Work

There are many works can be done in the future to improve the system. This first one is that edge detection can be used in detecting the location of the gate and fences so that the amount of manually inserted data by the user can be reduced. Next improvement that can be done to the system is that a model can be trained to identify the shape of the optical flow generated. Instead of relying on the vertical movement of the optical flow, the model will identify the shape of the flow and compared it to the shape of the flow generated by a jump over action.

Bibliography

Bodor, R., Jackson, B. and Papanikolopoulos, N. (2003). Vision-Based Human Tracking and Activity Recognition. *In Proceedings of the 11th Mediterranean Conference on Control and Automation*, 1, pp. 18–20.

Department of Statistics Malaysia, Official Portal 2019, *Crime Statistics*, *2019*,

https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=455&bul_id=MEs4QzNxWkNZZDEyM08yM0Jsd05vQT09&menu_id=U3VPMldoYUxzVzFaYmNkWXZteGduZz09 [Accessed 24 July 2020]

Eduardo A.B. d. S., Gelson V. M. 2005, *Digital Image Processing*, The Electrical Engineering Handbook.

Fonseca A., Mayron L., Socek D. and Marques O. (2008). Design and Implementation of An Optical Flow-based Autonomous Video Surveillance System. *Proc. IASTED*, 1, p. 209-214.

Fortun D., Bouthemy P. and Kervramm C. (2015). Optical Flow Modeling and Computation: A survey. *Computer Vision and Image Understanding*, 134, pp. 1-21

Lao W., Han J., De With P.H.N. (2009). Automatic video-based human motion analyzer for consumer surveillance system. *IEEE Trans Consumer Electron*, 55(2), pp. 591–598.

Mabrouk, A. B. and Zagrouba, E. (2018). Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert Systems With Applications*, 91, pp. 480–491.

Numbeo 2020, *Crime Index by Country 2020 Mid-Year*, https://www.numbeo.com/crime/rankings_by_country.jsp

[Accessed 24 July 2020]

OpenCV (n.d.). *OpenCV: Introduction* Available at: https://docs.opencv.org/4.3.0/d1/dfb/intro.html [Accessed 24 July 2020]

OpenCV (n.d.). *OpenCV: Optical Flow* Available at: https://docs.opencv.org/3.4/d4/dee/tutorial_optical_flow.html [Accessed 24 July 2020]

Ramachandran S., Palivela L.H., Vijayakumar V., Subramaniyaswamy V., Abawajy J. & Yang L. (2019). An intelligent system to detect human suspicious activity using deep neural networks. *Journal of Intelligent & Fuzzy Systems*, 36(5), pp. 4507–4518.

Poster

UNIVERSITY TUNKU ABDUL RAHMAN

Lim Kar Heng

Supervisor: Leung Kar Hang

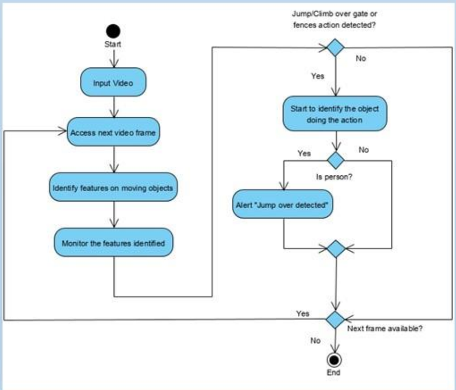
Introduction

Many families will install motion-detection sensors for security and surveillance systems. However, these systems have their own limitation. The alarm can be triggered by any moving objects such as a passing animal. A guard is required to constantly checking on the screen of what being recorded to know if a robbery is occurring for a normal surveillance system. A surveillance system using computer vision and pattern recognition will be better.

Objective

The main objective of this project is to develop a real-time surveillance system which will be better than the motion-detect sensor security or alarm system as this can reduce the false alarm triggered by any moving object. This real-time surveillance system also aims to reduce the delay in triggering an alarm. This is because the amount of human interaction required to trigger an alarm is reduced as it will automatically trigger alarm when a climb-over-gate action is detected.

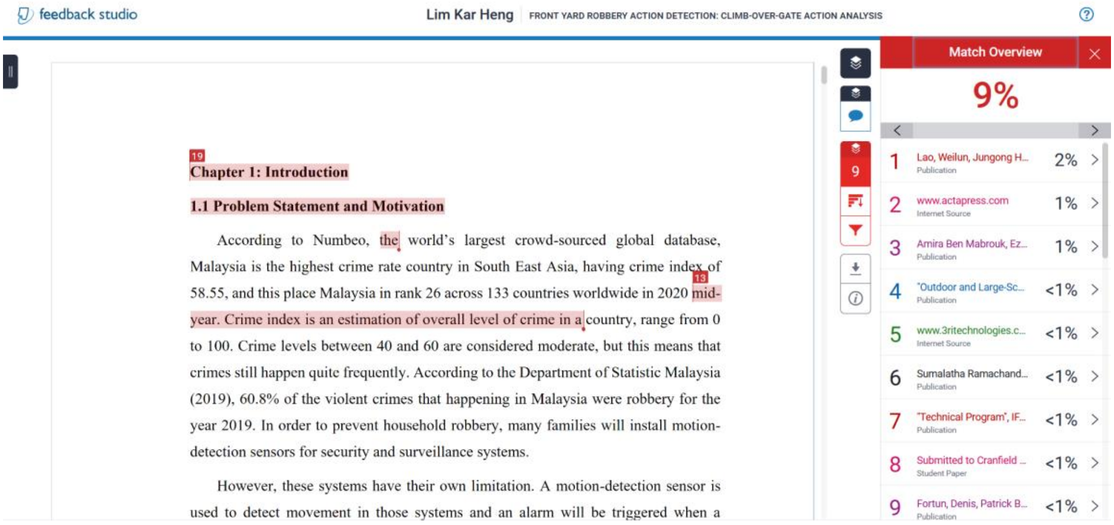
Methodology



Result

The result are shown below, when a person is climbing over the gate, an alert will be triggered.

Plagiarism Check Result



FRONT YARD ROBBERY ACTION DETECTION: CLIMB-OVER-GATE ACTION ANALYSIS

ORIGINALITY REPORT

9%	5%	6%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Lao, Weilun, Jungong Han, and Peter De With. "Automatic video-based human motion analyzer for consumer surveillance system", IEEE Transactions on Consumer Electronics, 2009. Publication	2%
2	www.actapress.com Internet Source	1%
3	Amira Ben Mabrouk, Ezzeddine Zagrouba. "Abnormal behavior recognition for intelligent video surveillance systems : a review", Expert Systems with Applications, 2017 Publication	1%
4	"Outdoor and Large-Scale Real-World Scene Analysis", Springer Science and Business	<1%

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	Lim Kar Heng
ID Number(s)	17ACB03158
Programme / Course	CS
Title of Final Year Project	Front Yard Robbery Action Detection: Climb-Over-Gate Action Analysis

Similarity	**Supervisor's Comments** **(Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index: __9__ **%** **Similarity by source** Internet Sources: __5__ % Publications: __6__ % Student Papers: __2__ %	
Number of individual sources listed of more than 3% similarity: _0_	
Parameters of originality required and limits approved by UTAR are as Follows: **(i) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.***	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

______________________________ ______________________________

Signature of Supervisor Signature of Co-Supervisor

Name: Leung Kar Hang Name: ______________________

Date: 16 Apr 2021 Date: ______________________

UNIVERSITI TUNKU ABDUL RAHMAN

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	17ACB03158
Student Name	Lim Kar Heng
Supervisor Name	Prof. Maylor Leung Kar Hang

TICK (√)	**DOCUMENT ITEMS** Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
/	Front Cover
/	Signed Report Status Declaration Form
/	Title Page
/	Signed form of the Declaration of Originality
/	Acknowledgement
/	Abstract
/	Table of Contents
/	List of Figures (if applicable)
/	List of Tables (if applicable)
	List of Symbols (if applicable)
/	List of Abbreviations (if applicable)
/	Chapters / Content
/	Bibliography (or References)

/	All references in bibliography are cited in the thesis, especially in the chapter of literature review
	Appendices (if applicable)
/	Poster
/	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report.	Supervisor verification. Report with incorrect format can get 5 mark (1 grade) reduction.
______________________	______________________
(Signature of Student)	(Signature of Supervisor)
Date: 15 April 2021	Date: 16 Apr 2021