FAKULTI KEJURUTERAAN ELEKTRIK

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

FINAL YEAR PROJECT 2

ANALYSIS OF PARTIAL OCCLUSION ON HUMAN PEDESTRIAN DETECTION

SASINTHIRAN A/L THEVENDRAN

B011510322

Bachelor of Mechatronics Engineering

May 2018

" I hereby declare that I have read through this report entitle **“Analysis of Partial Occlusion On Human Pedestrian Detection**" and found that it has comply the partial fulfilment for awarding the degree of Bachelor of Mechatronics Engineering”

Signature : ....

Supervisor's Name : ....

Date : ....

ANALYSIS OF PARTIAL OCCLUSION ON HUMAN PEDESTRIAN DETECTION

SASINTHIRAN A/L THEVENDRAN

A **report submitted in partial fulfilment of the requirements for the degree**

Of Bachelor of Mechatronics Engineering

Faculty of Electrical Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2018

"I declare that this report entitle" **Analysis Of Partial Occlusion On Human Pedestrian Detection** "is the result of my own research except as cited in the references. The report has not been accepted for any degree and is not concurrently submitted in candidature of any other degree"

Signature : ....

Name : ....

Date : ....

ACKNOWLEDGEMENT

First and foremost, I would like to express greatest thanks and gratitude to God for his guidance and blessings to complete this Final Year Project 2 successfully. I would also like to express my deepest gratitude to my advisor, Madam Nurul Fatiha Binti Johan, for her excellent guidance, caring and patience in providing me with suggestions, tips and encouragements throughout the completion of this Final Year Project 2. Not to forget my sincere thanks to Universiti Teknikal Malaysia Melaka (UTeM) that has facilitate me with laboratories and provide places for me to do researches and experiments.

Thanks to all my friends especially those who were always willing to help me by giving their best suggestions and encouragements. Special thanks to my parents for their non-stop support and love. Last but not least, thanks to everyone who are involved in the completion of my Final Year Project 2, directly or indirectly. However small your support is, it is still very helpful and meaningful to me.

ABSTRACT

Over the last few years, recognizing human in pedestrian surveillance system video is crucial for diverse application area. For an example, it can overcome the terrorism, some general social problems and violence by providing surveillance to the pedestrians. This also ensures to keep them in a very close watch for a secured environment. However, the main difficulty in human recognition is the occlusion of human and the light intensity. The common problems that influence the performance of human recognition system is the light intensity, distance of the camera, colour of the shirt or different poses of human. Occlusion can be happen in different ways, human with human or human with objects. In this system, a Webcam camera is used to analyse the recognition when the occlusion happens. Python software, OpenCV library and Keras API are used to train the system to recognize the human. Haar Cascade is used to track humans and the Convolutional Neural Network is used to classify the human models. The webcam camera captures the image and recognizes the human that passes by it. The data is taken from the real time video of the system and is used to analyse the accuracy and performance of the system. The analysis proved that this system has high accuracy in recognizing human which is 97.35% of accuracy. It only has 2.65% rate of error. The occlusion analysis shows that the accuracy decreases as the occlusion increases. Then, the results clearly shows that moderate lighting is the perfect lighting for this system and when the distance increases the accuracy decreases. The distance problem maybe caused by the ability of the camera. So, this system performs very well in human recognition and it also has high accuracy.

ABSTRAK

Kebelakangan ini, pengesanan manusia dalam video sistem pengawasan pejalan kaki amat penting untuk pelbagai aplikasi. Sebagai contoh, ia dapat mengatasi keganasan dan masalah sosial dengan mengawas pejalan kaki. Ini juga memastikan mereka sentiasa berada di bawah pantauan untuk persekitaran yang terjamin. Walaubagaimanapun, kesukaran utama dalam pengesanan manusia adalah oklusi manusia dan keamatan cahaya. Masalah umum yang mempengaruhi prestasi sistem pengesanan manusia adalah keamatan cahaya, jarak kamera, warna pakaian atau pose manusia yang berbeza. Oklusi boleh berlaku dengan pelbagai cara yang berbeza, manusia dengan manusia atau manusia dengan objek. Tujuan projek ini adalah untuk mengesan pejalan kaki dengan menggunakan kaedah Haar Cascade dan juga kaedah Convolutional Neural Network. Kemudian, objektif kedua sistem ini adalah untuk menganalisis prestasi sistem pengesanan manusia apabila berlakunya oklusi. Dalam sistem ini, kamera Webcam digunakan untuk menganalisis pengesanan manusia apabila oklusi berlaku. Perisian Python, OpenCV library dan Keras API digunakan untuk melatih sistem untuk mengesan manusia. Haar Cascade digunakan untuk mengesan manusia dan Convolutional Neural Network digunakan untuk mengklasifikasikan model manusia. Kamera webcam merakam imej dan mengenali manusia yang melaluinya. Data diambil dari video masa sebenar sistem dan digunakan untuk menganalisis ketepatan dan prestasi sistem. Analisis ini membuktikan bahawa sistem ini mempunyai ketepatan yang tinggi dalam mengiktiraf manusia iaitu sebanyak 97.35%. Ia hanya mempunyai kadar ralat sebanyak 2.65%. Analisis oklusi menunjukkan bahawa ketepatan semakin berkurang apabila oklusi semakin bertambah. Kemudian, hasilnya dengan jelas menunjukkan bahawa pencahayaan sederhana adalah pencahayaan yang sempurna untuk sistem ini dan apabila jarak kamera meningkat, ketepatannya menurun. Masalah jarak kamera mungkin disebabkan oleh keupayaan kamera. Oleh itu, sistem ini berfungsi dengan baik dalam mengesan manusia dan ia juga mempunyai ketepatan yang tinggi.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

γ	= Wavelength
C	= Speed Of The Light
E	= Energy
F	= Frequency
H	= Planck 's constant
$ii(x, y)$	= Integral Image
$\mathrm{i}(\mathrm{x}', \mathrm{y}')$	= Original Image
$s(x, y)$	= Cumulative Row Sum
X	= Input Vector
H	= Vector Of N Neurons
Y	= Output Vector
w_{ki}	= Weight
w'_{ij}	= Weight between i^{th} Hidden Layer And j^{th} Output Layer
W1-Wk	= Weight Vector
F	=Nonlinear Function
Y	= Scalar Output (Prediction)
F	=Takes A Weighted Sum

LIST OF ABBREVIATION

CNN = Convolutional Neural Network

SVM = Support Vector Machine

ROI = Region Of Interest

HOG = Histogram Oriented Gradient

DPM = Deformable Part Model (DPM)

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

API = Application Programming Interface

CHAPTER 1

INTRODUCTION

1.1 Motivation

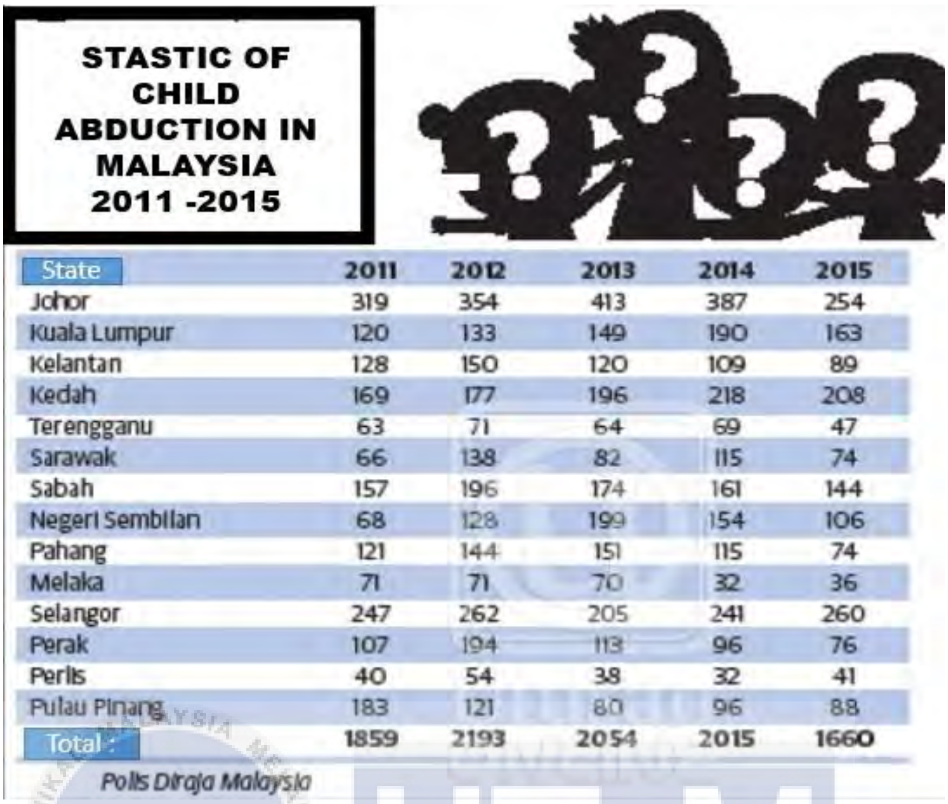
A pedestrian is a person who travels on the road either by walking or running. Nowadays, the highest rate of crimes happen mostly to pedestrians walking on the road especially at night. Such crimes include murder, robbery, theft and rape. Thus, the pedestrians themselves feel very insecure about their safety out there.

Based on a recent research, it was clearly stated that every year, more than 270000 [1] pedestrians become prey for such crimes mentioned above. News about these crimes are always being flashed in the newspaper, television and also social media daily. Based on the PDRM statistics, it can be clearly proven that these crimes are becoming more popular than before. The gang related crimes are becoming a trend nowadays and is increasing day by day. Besides, the pedestrians themselves also break the law of the roads. This leads to losing their valuable life. The carelessness of the people also affect the safety of others. Only few cases have been caught by the police but there are still many without the right proves. Figure 1.1 shows some cases that the police have caught.

TYPE OF CRIMES EXPERIENCED	2012 to 2014 %	2005 to 2011 %
House theft	42.4	42.7
Snatch theft	76.4	70.8
Car theft	21.5	26.0
Theft in a taxi	9.0	4.2
Other theft	16.0	20.8
Physical assault	27.8	21.9
Rape	2.1	1.0
Other crime	16.0	8.3
Note: Only for those with direct experience of a crime		

Figure 1. 1: Statistic of crime cases in Malaysia 2005 – 2014 [1]

Apart of that, the children and women pedestrians are also becoming a main target for such crimes. Based on a statistic, it can be said that the child abduction is increasing drastically every year even when their parents are with them. The women pedestrians also feel very insecure about their safety while walking alone somewhere because of some criminals that take advantage of the women's weakness. Some of the victims of these cases have lost their life due to bad health condition after the abuse and also some by committing suicide. Figure 1.2 shows the statistic of children abduction in Malaysia.



State	2011	2012	2013	2014	2015
Johor	319	354	413	387	254
Kuala Lumpur	120	133	149	190	163
Kelantan	128	150	120	109	89
Kedah	169	177	196	218	208
Terengganu	63	71	64	69	47
Sarawak	66	138	82	115	74
Sabah	157	196	174	161	144
Negeri Sembilan	68	128	199	154	106
Pahang	121	144	151	115	74
Melaka	71	71	70	32	36
Selangor	247	262	205	241	260
Perak	107	194	113	96	76
Perlis	40	54	38	32	41
Pulau Pinang	183	121	80	96	88
Total :	1859	2193	2054	2015	1660

Figure 1. 2 : Statistic of child abduction in Malaysia [2]

So the researchers have found a solution to catch the criminals and also to reduce the rate of crimes in Malaysia at which the solution is “Human pedestrian recognition for surveillance system”. The researchers feels that it can reduce the crimes and also help the police by providing them with evidence to sentence the criminal. The most important outcome of this system is that the pedestrian themselves will feel very secure from the crimes.

All these points mentioned above shows the importance of the “pedestrian recognition system”. It also gives us prioritize in introducing this new system. This system surely can reduce the criminal activities happening to the pedestrians on the road. It can also be an effective way to reduce the feel of insecurity of pedestrians. To overcome the crime activity that being spread on the pedestrian walk or can be said the spread of the street crime, maybe this human recognition system can be the most suitable system to be used and this system also will improve the features of the current surveillance system.

1.2 Introduction

Computer vision [3] is a field that focuses on how computers can understand the images or videos that are being recorded by the camera. The previous researches about image processing show the difficulties faced in the computer vision field. So currently, the researchers are mainly focusing on how to develop the computer vision field to become easier. There is a close relationship between the computer vision and pedestrian recognition system which is the image processing. The usage of the real time pedestrian recognition system has increased nowadays and it has also become a new innovative development for mankind. Besides, it is also a useful tool for research in detecting, tracking, analysing and counting the behaviour of an object in the video. There are a lot of variance that can influence the human recognition system. The researchers are still finding many ways to make the recognition system works easier and faster. The video and image analysis is becoming more popular and have high demand in market because of its wide applications with rapid growth in the technology of camera. This system has a new section for human recognition and also will help to track down the human. The human recognition system can be a way to analyse human behaviours, to count humans and the most important thing is it can be used for security purposes. Figure 1.3 shows an example on how human is detected when they pass by the camera.

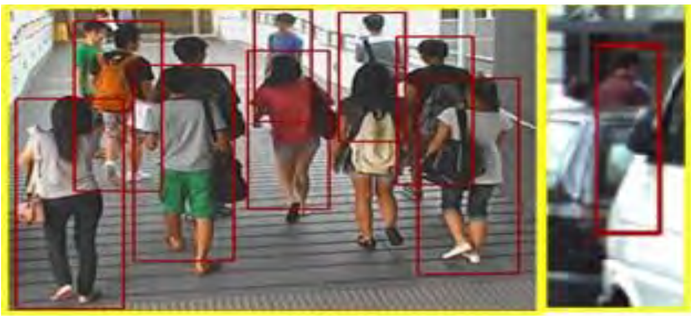


Figure 1. 3 : Human pedestrian detection system [5]

The basic parts in this system are object detection, object segmentation and object recognition. Object detection is the process of finding entities of some real-world objects such as humans, faces, cars in the form of images or videos. This pedestrian recognition system should detect humans with a bounding box around each of them.

Object segmentation is also one of the part in image processing to build this system. Object segmentation is the process of extracting the target object from the videos or image sequences. The object that we want to focus is the human and the background of the image, therefore the segmentation should highlight the pedestrian in this system.

Lastly, object recognition is the most important part in this system. Object recognition is a very difficult part in computer vision and it requires more data to train or teach the computer. More time is required to train and analyse the computer. This system must detect the object and recognize it by comparing the object with the database whether it is person A or person B according to their face and body parts.

1.3 Problem Statement

In today's world, the real time human pedestrian recognition system plays an important role in pedestrian safety purpose. The recognition system has the ability to recognize a person and store the information into the database for future usage. Each and every task in this human recognition system takes place automatically. This recognition system can be used for various applications such as sign board recognition, safety surveillance system, car manufacturing and etc. The recognition system actually works by enabling a computer to identify an object. The computer originally sees the objects in a form of big matrix of pixels and numbers of representations. Humans can easily classify any objects from an image or video through their eyes whereas the computers have no idea with what is in an image or video because it can only see numbers and matrices. Hence, the computer requires more data to calculate and extract the features so that the object can be identified.

Besides, to enable a computer to identify an object such as human, there are many problems to be faced. The very first problem is the viewpoint difference, which is the difference in heights of the humans that makes the computer less precise when it comes to recognition. Secondly, the distance between the camera and the human also influences the detection process where if the pedestrian is not within the range of detection, it might affect the camera detection. The third problem is the light intensity. It is also one of the biggest challenges in this system. Different lightings and environments increase the difficulty in detection of the camera to track human. Besides, human wear is also a problem in this system. When the human wears different types of clothes, colours and styles, it is hard to find the universal representation of a human. Apart from that, the recognition system must be able to handle the occlusion of humans as well. Occlusion can be happen in different ways, human with human or human with objects. So, Occlusion might affect the accuracy of the system.

Therefore, this recognition system requires more data of humans to improve the accuracy of the system as well as to handle the problems mentioned above. The higher the number of data, the higher the accuracy of the computer to learn and recognize human.

1.4 Objectives

This project will embark on the following objectives:

1. To recognize the human pedestrian by using Haar Cascade and Convolutional Neural Network methods.
2. To analyse the performance of the human recognition system when occlusion happens.

1.5 Scope and Limitation

This project is mainly focuses on indoor pedestrian walk and twenty person's datasets are used for the training and testing of the system. The main parts in this system is detection and recognition. The webcam camera records the real time video and sends the recording data to the computer to display the video stream. The frame rate of the video stream is set constant at about 60 frame per second (FPS) by adjusting the coding. Besides, the computer vision techniques are done by using the Python software which is the easiest language. Then, the Haar Cascade and Convolutional Neural Network by Keras API methods are used to classify the features. Three experiments are conducted in this system by using twenty pedestrians and at four different places. The analysis or experiments are mainly focused on the performance of the system when the occlusion happens.

CHAPTER 2

LITERATURE REVIEW

In this section, several studies have been conducted regarding "Human Pedestrian Detection and Recognition System". The flow of studies started from techniques used, software used, effective methods used and algorithms used to detect and recognize the human in the real time video stream. Many studies have been conducted to find the best technique to detect and recognize human. This chapter will explain about all the studies from previous researches and also analyse their comparisons to find a very suitable method for this system.

2.1 Theoretical Background

Human recognition is the most fundamental task in this project. This chapter will introduce several works and techniques that are related to the human pedestrian detection and recognition. There are several researches on successful detection and recognition of human and various methods are used to perform these detection tasks. This project mainly focuses on human pedestrian recognition which needs to be performed in real time video stream. The images or videos that have been captured by camera (vision system) must be able to detect human pedestrians and recognize each of those whom are passing by the camera. As all knowns, human exist in multiple size, heights and colours. In order to develop a system to recognize human, first step will be finding all the techniques and understanding the methods involved. So, all the researches are performed based on human recognition.

The fundamental knowledge that is needed to ensure the succession of project is the knowledge of machine vision and deep learning. Machine vision focuses on image processing whereas deep learning focuses more on artificial intelligence. Both studies are very important for this project because the methods in those studies will be used for this project to recognize human. Researches are needed when choosing suitable methods, software and libraries because the best selections can lead to the succession of the project and always works around the scope of the project. Besides, the data and information that are gathered from the previous researches might be able to improve the understanding in methods and techniques used and also able to find out the problems and solutions that the previous researchers have handled. Hence, from the studies conducted, this project can be prevented from any kind of problems that have been faced and handled by the previous researchers and also will be a way to improve the knowledge in human recognition system.

2.1.1 Machine Vision

Machine vision is a system which uses computer to sense an object through camera or a visual sensor. It has the ability to extract information automatically from the image of the real scene. Machine vision uses the images or videos captured and analyses it for automating tasks such as image manipulation, detection, tracking, inspecting and counting. Today, machine vision can reach consumers in many contexts such as via webcams, camera phones and sensors. It provides high level of interfaces for capturing, pre-processing and presenting image data. There are two main things need to be considered in machine vision which is the sensitivity and resolution [5]. Sensitivity is the ability of a machine to see objects in different lights and detect any objects with weak impulses at invisible wavelengths. Resolution is the extent to which a machine can differentiate between objects. In general, the better the resolution, the more confined the field of vision. If two images are used to do pre-processing in machine vision at which one of the image has a resolution of 900x900 pixels and another one with a resolution of 300x300 pixels, the image with 900x 900 pixels will be more accurate to do pre-processing with machine vision system. Highest Random Memory (RAM) is needed for high resolution camera and Artificial Intelligence (AI) program is also needed to operate machine vision system. The tools used in machine vision system is digital or analogue cameras, digitizing images and a processor. In this machine vision system, the computer programs will perform several tasks depending on its requirements and the images and videos will be analysed by this system automatically. Figure 2.1 shows the example of machine vision system.

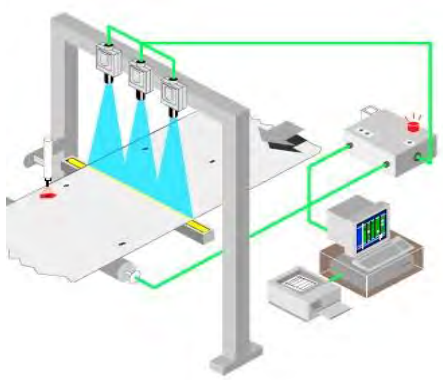


Figure 2. 1: The computer vision system [5]

2.1.2 Deep Learning

Deep learning is a subfield of machine learning, which is in turn, a subfield of Artificial Intelligence (AI). The goal of AI is to provide a set of algorithms and techniques that can be used to solve problems that humans perform intuitively and automatically, but are otherwise very challenging for computers. As an example of AI problem is interpreting and understanding the contents of an image. This task is something that a human can do with little or no effort but it has been proven to be extremely difficult for machines to understand the contents of an image.

The word deep in deep learning does not mean any kind of deep understanding achieved by the approach but it's about the representations of the layers [7]. The number of layers contributed to a model of data is called the depth of the model. These layers have the capability to learn the input data automatically. These layered representations are learned via models called neural networks. Neural network is a reference to neurobiology and is developed in part by drawing inspiration from the understanding of the brain. So, deep learning is a mathematical framework for learning representations from the data. Figure 2.2 shows the architecture of deep learning.

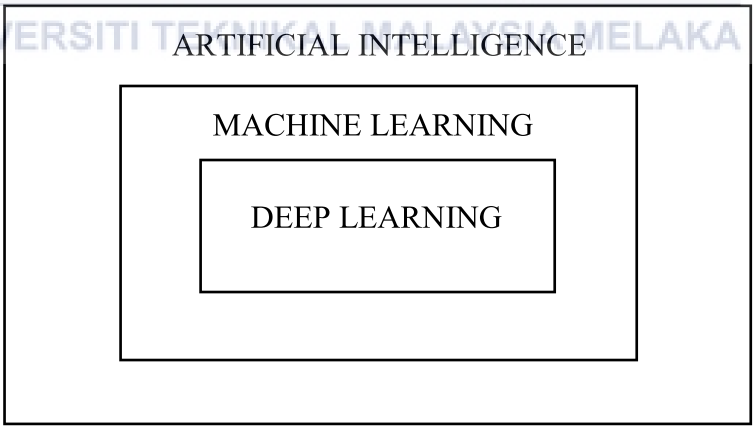


Figure 2. 1: The Architecture of deep learning

2.2 Overview of previous study

2.2.1 Human detection system

This topic will explain about all the human detection systems that has been designed by the researchers. It also will explain all the methods and techniques used for the detection system.

2.2.1.1 Real time human detection using background subtraction method

A real time computer vision system is mainly used for pedestrian detection, human tracking and verification when challenging situation occurs. There are some algorithms that are used to complete this system which is the machine vision algorithm, human tracking algorithm and motion analysis algorithm [8]. The main objective of this system is to improve the robustness and efficiency when detecting human in real time videos. This system will integrate all the algorithms to increase the efficiency of the detection. The human detection system is used to decide whether an image contains a human or not. The tracking algorithm uses the intensity to detect human over time. The motion analysis algorithm is used to analyse the motion to identify whether the detected and tracked objects works like a human or not. For the computational calculation and real time performance, this system uses some computational resources such as multi-threading, high performance library and efficient inter-thread communication. Multi treading will clear the unnecessary blocking of the system. Three modules are used for these three algorithms which are the detection module, object tracking module and the motion analysis module. The advantage of this system is the ability in detection of the pedestrians even when the environment is in uncontrolled camera motion. Figure 2.3 shows the steps involved to complete this real time human detection system.

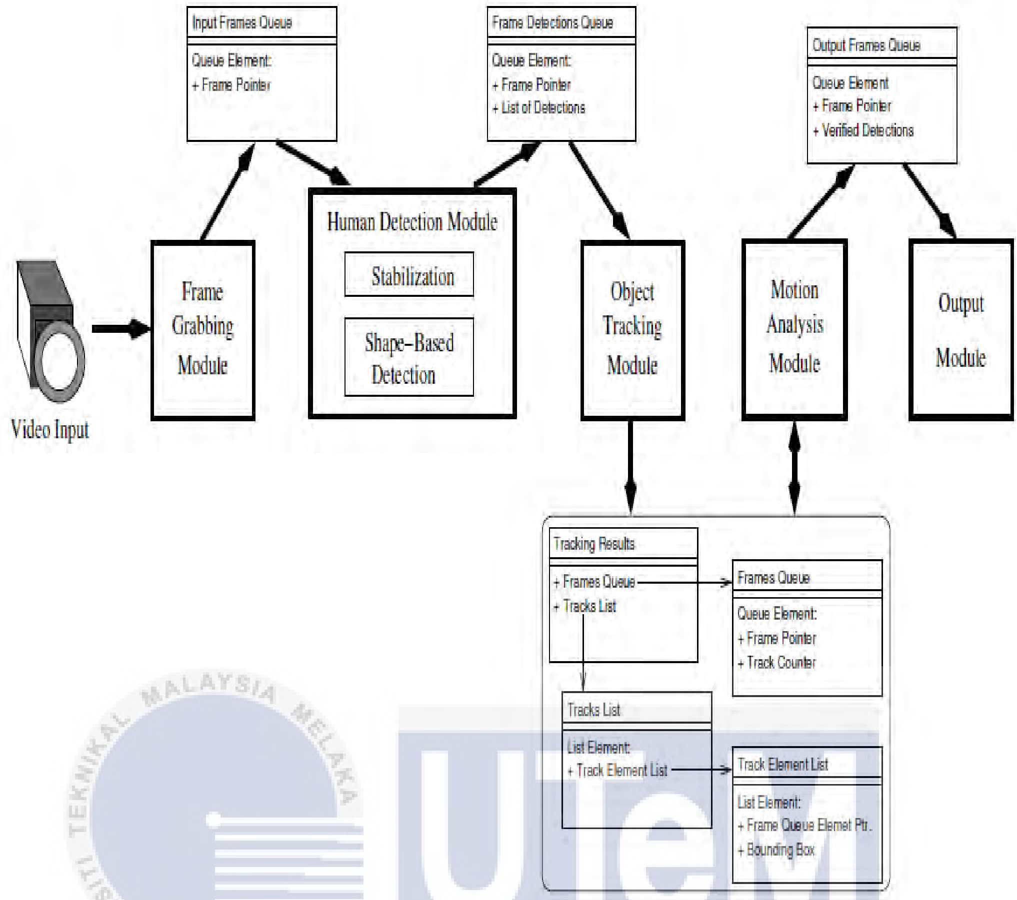


Figure 2. 2: The architecture of the real time human detection system [8]

2.2.1.2 Occlusion and abandoned object detection by Gaussian technique

Occlusion can be divided into three types. First is self-occlusion, the second one is inter object occlusion and the third one is background occlusion. The aim of this system is to develop a framework to detect and track the pedestrians even when heavy occlusion happens. For abandoned object detection, some steps have been applied which is differencing the image, average intensity calculation and Support Vector Machine classifier. There are some algorithms that have been used for this system which is the Background Subtraction, blob detection and Histogram Oriented Gradient. The background subtraction is executed by a mixture of Gaussian technique which will detect and track all the motions of human. The components that are grouped in the background subtraction are background modelling, threshold selection and pixel classification. After the background subtraction, each object is detected as a blob. Each blob will perform Gaussian background subtraction and basic filtering. The blob will filter the contents of the image and is executed by the OpenCV Library. Then, the Histogram Oriented Gradient will detect the human by using the Support Vector Classifier and it is also used to count the existence of gradient orientation in limiting the dataset of the picture. The differencing image is a result from filtering the scenes at two continuous time instant from each other.

Figure 2. 3: The flow chart of this object detection system [9]

2.2.1.3 Human tracking in video surveillance with temporal differencing

The human tracking system is used only to detect humans in real time video surveillance. This system is very effective in detection and is used to handle the occlusion using velocity and information of the direction. Basically, there are three approaches in moving human detection which is temporal differencing, background estimation and optical flow methods. Temporal differencing [10] is modifying the dynamic environments. Optical flow is used to detect moving objects with the presence of the camera motion. The background subtraction is used to detect moving region in a picture. Then, for a real time video surveillance system, some methods need to be applied such as image separation, human detection, noise removal, human tracking and analysing the activity. Image separation process is used to divide a video into a sequence of images. Then, the noise removal need to be proceeded before the image is sent to the following processes because some part in a video might be added and erased when obtaining images. There are two methods that can be used to track objects which is the region based tracking and contour based tracking. By using those methods, this system is able to track human in real time videos.

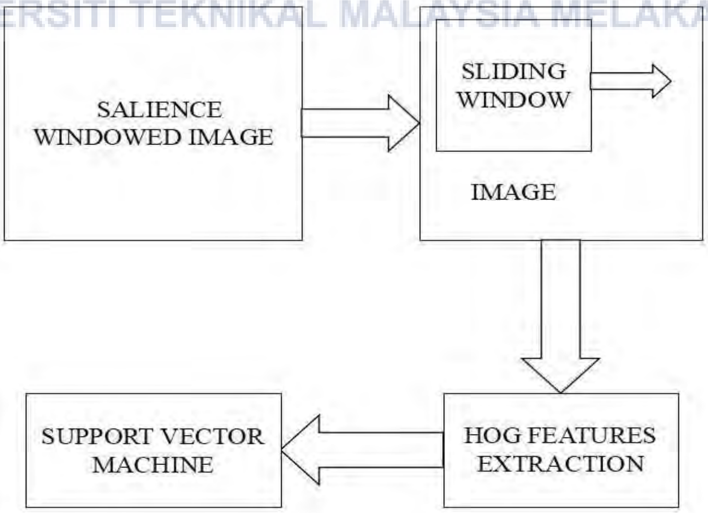


Figure 2. 4: The flow of the human tracking system [10]

2.2.1.4 Human motion detection and tracking by Histogram Oriented Gradient algorithm

Motion is one of the key point to detect human pedestrians easily nowadays. This human tracking system also uses the motion to track objects and identify which one is human and non-human. The tools that are used in this system is Background modelling algorithm and Histogram Oriented Gradient algorithm [11]. This system captures several videos to find the accuracy of detection system. From the video scenes, the contents of the image is extracted and eliminated from the original frame. The background model reads only the static parts of the scene and eliminate the moving parts. This system uses the Gaussian Mixture Model for the background modelling and it uses the Region of Interest (ROI) to identify the motion entities. The ROI is a dataset which consists of human models, animals and vehicles. The Histogram Oriented Gradient algorithm is used in ROI to identify which category of object is present. Then, the image will separate into multiple entities of fixed rectangle size. The hog will calculate each cell in the rectangle size. Besides, the local contrast normalization is needed to reduce the problems of the illumination variation in the image. The light support vector is used to teach the feature vector in this tracking system. The advantage of this system is it has the ability to classify an object whether it is a human or non-human. The disadvantage of this system is it can only detect one object at a time.

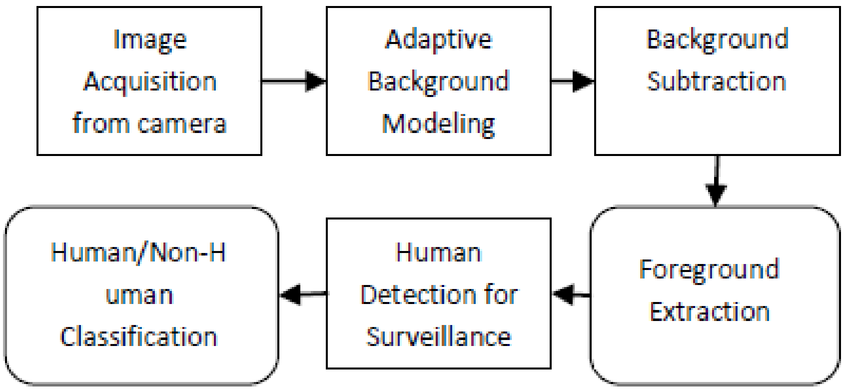


Figure 2. 5: The overview of human motion tracking system [11]

2.2.1.5 Real-time people counting system by curve analysis method

This system for real-time counting people in a video or image is more effective and flexible in counting. This system uses the image processing technique. The main problem issued in this system is on how to filter the essential data from an image and count people accurately by using Curve Analysis method [12]. There are two steps in this system which is image processing and image understanding. The image processing will separate the object needed from the background and other objects. Then, the image understanding is applied by the features of understanding technique. The template matching is applied to count people from an image. Monochrome images are used in this system to eliminate the light intensity. This system change the image into binary image which is the object used in the image is declared by black pixel and the background is declared by white pixel to eliminate the object from the background. The resolution of the image must be 318x238 in pixels. Besides, the segmentation technique is used to filter the object out from the background. The counter of the image is separated into sub blocks. The edge detector Sobel filter is used to avoid the shadow corruption. Then, the median filter is used to eliminate noise. Besides, the curve analysis method is developed to count human in an image after the segmentation process. The cluster segmentation is used in this system. The advantage of this system is it has the ability to count people accurately.

2.2.1.6 Motion detection and tracking using background subtraction

The camera of this video surveillance device detects a human motion and then processes the motion detection and conducts warning alarm. There are two methods used in this system which are the background subtraction and frame difference methods [13]. The Background subtraction method is used to detect the moving object by counting the difference between the original frame and a reference frame for this system. By using a fixed camera, this system is able to detect motions and compare the original image with the background image. It will compare pixel by pixel and calculate the number of changes in pixel. If the pixel number is more than threshold number, it means motion is detected. Then, it will subtract the object in the image from background. A graphical method is also used in this system to show the number of the detected objects and the percentage of detection. The advantage of this system is it has the ability to detect moving objects accurately.

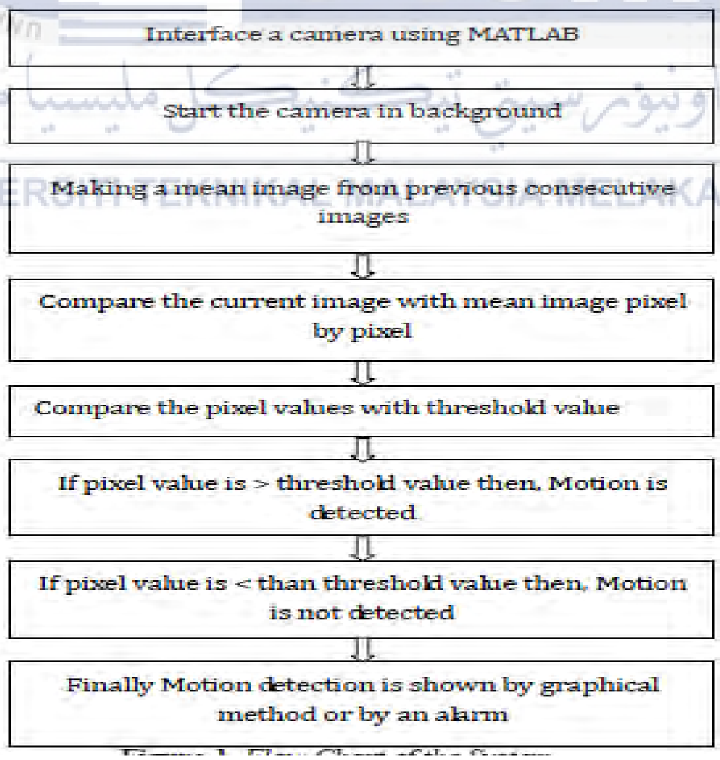


Figure 2. 6: The flow chart of this human tracking alarm system [13]

2.2.1.7 Summary of Previous Studies (Computer Vision)

Based on the previous studies, many researchers have used the Histogram Oriented Gradient (HOG) technique as a popular method for object detection. The HOG technique have shown good results in human detection systems. It also has no false detection error of human pedestrian. Some of the researchers have used other methods such as using Background Subtraction, Deformable Part Model (DPM), Temporal Difference Technique and Haar Cascades as detecting techniques. The Haar Cascade technique also performs very well and is easy to use in human detection system. Following will be a brief explanation of each technique.

Table 2. 1 : The comparison of machine vision methods

Method	Description	Advantages	Disadvantages
Histogram Oriented Gradient (HOG) [14]	Histogram of Oriented Gradient (HOG) is a feature analyser which is widely used to identify object through the object's shape. The method was introduced by Dalal and Triggs. This method is mainly used in computer vision and image processing to detect an object such as human face / human body. There are some steps in HOG detection. First, it will divide the original image into blocks. Then, each blocks will be separated into smaller regions called cells. Normally, the block lies on each other. So the cells will be in several blocks. From the cells, each pixel will be taken to obtain the vertical and horizontal gradient. Then, hog will be created in each of the cells. To extract the features from the hog, the Support Vector Machine (SVM) algorithm is used.	• Able to extract global features • Able to detect human • Able to track human	• Works well for smaller resolution • Requires more complex reasoning

Method	Description	Advantages	Disadvantages
Support Vector Machine (SVM) [15]	Support Vector Machine (SVM) is an algorithm which is used to classify an object. It is also mainly used in regression analysis. Some researchers use the SVM as a single layer neural network. The SVM classifier makes a new model to assume the classes for new examples. There are two types of SVM classifier which is Linear SVM classifier and Non Linear SVM classifier. The SVM will classify the type of object that has been extracted from the feature of the HOG.	• Ability to solve computational problems • It can avoid over fitting in image processing	• It has the limitation in speed and size • Support vector approach lies in choice of the kernel
Background Subtraction [15]	Many researchers mainly focus on this method which is also known as Foreground extraction because it has the ability to extract a foreground frame and image. The background subtraction is used to count the foreground image working on a subtraction between the current frame and a background part. This method enables it to extract the scene from the video or images. It can also detect the moving particles by comparing the current frame and background image.	• The object is taken from the background of image without destroying the quality.	• Cannot deal with lighting changes

Method	Description	Advantages	Disadvantages
Deformable Part Model (DPM) [15]	Deformable Part Model (DPM) is used to train part based model which is the local humans appearance template and the dataset of spatial connections. It has the ability to detect moving objects and human. In this method, the training data will be shared with the deformations. This method is also mainly used in human detection.	• Perform high accuracy and speed in detection	• It has large amount of calculation
Temporal Difference Technique [16]	Temporal Difference technique is defined as learning about the prediction of the quantity that depends on upcoming values of a given value. In image processing, it is used to predict the image based on the machine learning method. The prediction will adjust the observation to be better.	• Temporal difference learning is faster in detection	• It has less stability in detection
Haar Cascade [16]	Haar wavelet is a systemic way to perform an image processing. It is a square shaped function and it works as the Fourier analysis. It uses the average and difference value in image to form a matrix. In human detection, this technique purposed a wavelet scale for representation in an image. The scale of the wavelet will be about 16 or 32 pixels.	• Accurately analyse the signal • Very easy to train the cascade model	• The signal is not continuous

2.2.2 Human Recognition System (Deep Learning)

This topic will explain about all the human recognition systems that has been designed by the researchers before this. It also will explain all the methods and techniques used for the recognition system by using the deep learning method.

2.2.2.1 Real-time pedestrian detection with deep network cascades

This system introduces a new real-time approach that improves the efficiency of cascade classifier of deep neural networks. The objective of this system is to produce a fast and accurate detection system for pedestrians. The approaches used in this system are cascades deep nets and fast feature classifier [17]. This system uses 'Doppia' open source as an algorithm. Then, a fast feature classifier is added with a cascade of deep neural network to make this system detect humans in a fast way. Besides, the deep network have only three hidden layers, which is 5x5 convolution, fully connected and 1x1 convolution layer. The layers are used to speed up the system and also used to filter the contents in the video. It works 80 times faster when it is combined with two deep neural network. This system works as follows which at first, a square box on pedestrian examples is cropped. Then, the image is resized to 72x72. Besides, it also saves the hard negatives. Then, the pedestrians are extracted from the image. An algorithm is used to convert the video stream in real time at 60 frames/second. The advantage of this system is that it has the ability to detect an object in high speed and very accurate.

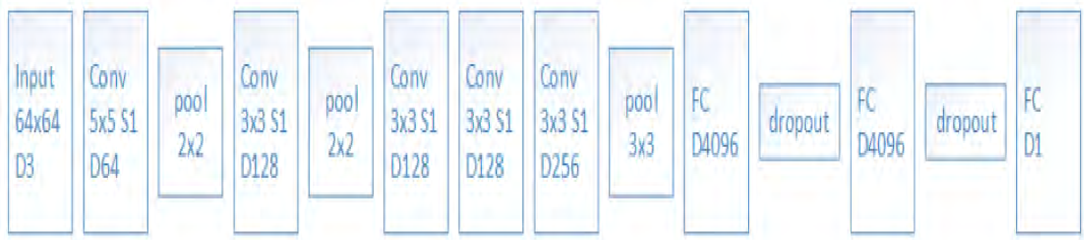


Figure 2. 7: The architecture of deep neural network [17]

2.2.2.2 Human detection through processing video captured by camera module interfaced using raspberry and OpenCV

This is a simple hardware model of human detection system using Raspberry Pi, which is a mini controller. The Raspberry Pi works as a controller which transfers the video to the screen connected. The coding are performed in python language. This system is capable for motion detection and human tracking. This system also has the ability to recognize humans, calculate the number of humans and make a bounding box around the human. First step of this system is it captures images and videos by using a Pi camera and transfers the captured data for pre-processing. The pre-process is done by the OpenCV library. The pre- processes are resizing the image, filtering and labelling of the image. For this system, Viola Jones [18] algorithm is used to train the system to detect an object. The algorithms are performed in image segmentation, localization and extracting the feature. This algorithm also has four stages. The first stage is selecting the floor feature, second is generating integral image, then Adaboost training and the last is cascade classifier. The advantage of this system is cheap, small and compact but it has high performance in detecting people in real time.

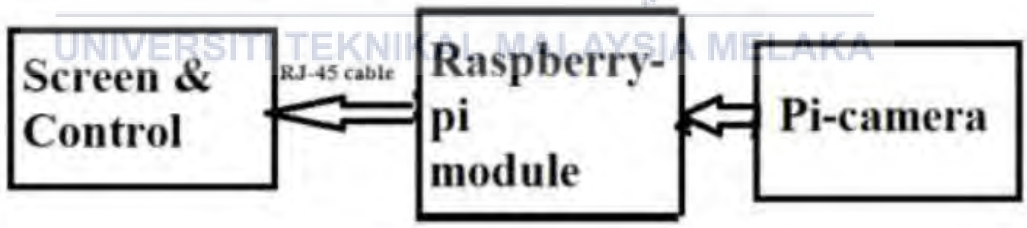
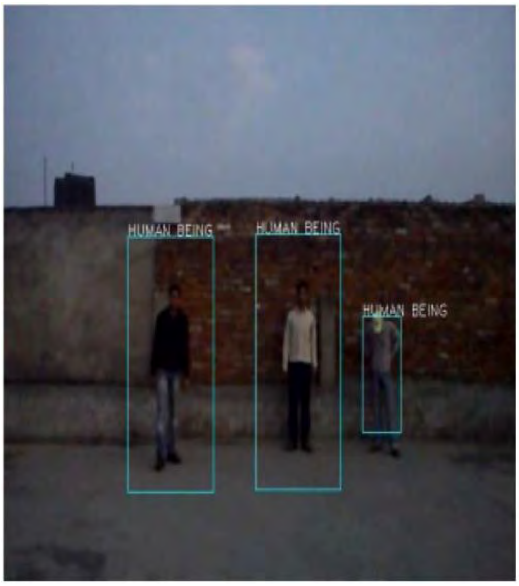


Figure 2. 8 : Block diagram of the Raspberry Pi based human detection system [18]

2.2.2.3 Automatic multiple human detection and tracking by Haar like Features

It is an object tracking system which is able to detect multiple objects such as humans in outdoor environment. In this system, Haar-like features [19] work as trained classifier and Viola Jones works as framework hich that is used for detection of human. Humans and objects can be tracked by this trained model of Haar features and also by using some particle filter. The machine learning is used to create the real time video stream and to speed up the classifier. Once this system detects human, the next step is to track the visual of human. There are two methods to track multiple objects by using particle filters which is creating multiple particle filter and using a single particle filter. The single particle filter works well in tracking if it is not occluded by anything. In this system, some datasets of humans are taken for training the detector to perform well in tracking multiple human even in a critical lighting conditions. The Haar-like features will perform the calculation of this system. There are some steps to filter the contents from the videos, which is first, predicting the human, second resampling the particles by the weight, third is creating a model and the last step is completing the algorithm. The result shows that this system has achieve 87.44% accuracy in detection and the average speed of processing is 18fps.



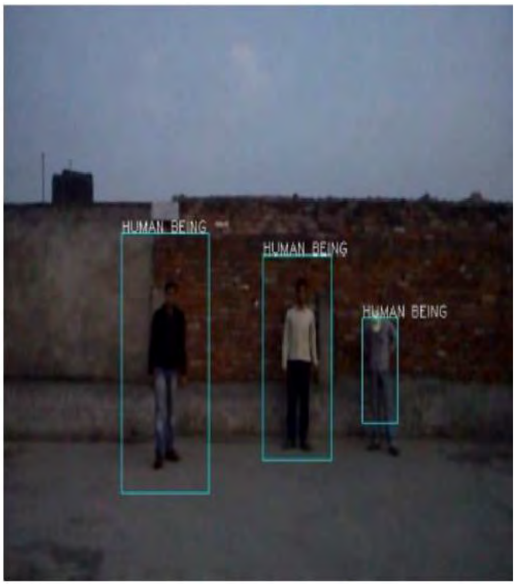


Figure 2. 9 : The result of this human detection system [19]

2.2.2.4 Object classifications technique using machine learning

This system aims to show experimental study of the human classifications. In this system, Support Vector Machine (SVM) algorithm is used as a classifier and 4000 pedestrian datasets has been collected to train the dataset. The artificial system is used to extract the information from the images and videos. The main problem that has been faced in this system is object detection, classification and tracking. This computer vision system will be followed by some tools which are data acquisition, the representation of data and a decision part. The data acquisition [20] uses the digital image processing techniques. Second, the representation of data will extract the required information from the videos. Then, the third is decision part which is it will recognize the parts. There are four steps used to develop the classifier in this system which is create a training set, select a feature, train a classifier and improve the classifier efficiency. There are some steps used to complete this project. First step is to collect the datasets of pedestrian which will be used as input. Second is finding a feature which uses machine learning and computer vision to perform pre-processing tasks. Third is transformation of the features. Then the last step is selected features of the images will be divided into two datasets which is training data and test data. The advantage of this system is it has the ability to recognize human.

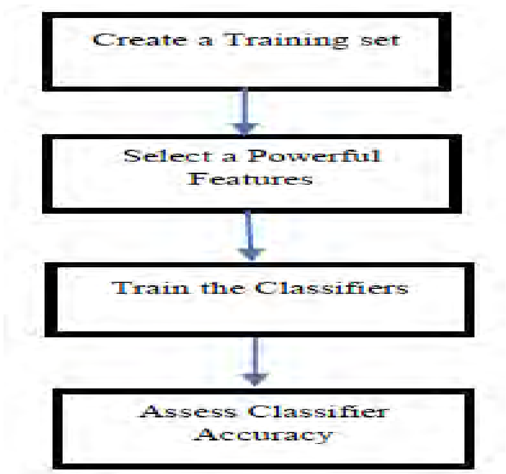


Figure 2. 10: The steps of classification of this human recognition system [20]

2.2.2.5 Human activity recognition with Convolutional Neural Network

This system aims to recognize the human activity by using combination of deep networks. There are two networks in this system which is modified weber descriptor and local pattern descriptor. The modified weber descriptor is used to extract features from image and the local pattern descriptor is used to extract the information of spacial of human body. The extracted features will be moved to the convolutional Neural Network (CNN) classifier. It is used to differentiate the object's activity. To train this system, two dataset samples have been taken which is HMDB and Hollywood dataset. Interest points are given some information on the body movement. The steps for completion of this system is first, a video is analysed frame by frame. Then, the humans are detected by using the HOG method. To calculate the local features of the human, this system uses the SVM algorithm. Then the features are collected to make calculations and predictions. The output from the dual networks are moved to CNN classifier. The interest point is mainly used in this system which is skeletal image of the human to extract information from the videos.

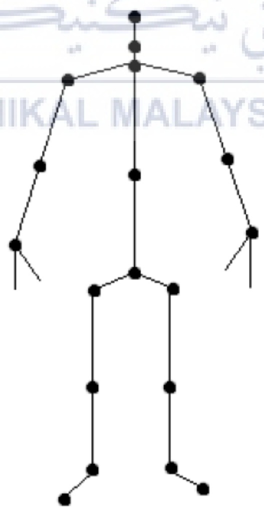


Figure 2. 11: The interest point of human [21]

2.2.2.6 Summary of previous studies (Deep Learning)

Based on previous machine learning studies, many researchers used deep neural network for human recognition because it is the fastest and accurate way. Many researchers focused on the CNN network which integrates by few layers to classify the humans or objects from an image and videos easily. Some of the researchers are using Haar-like features and Viola Jones method as a classification tool. Then, the researchers used some libraries such as Tensorflow, Caffe and Teano to perform the Deep Learning in human detection system. Comparisons as below:

Table 2. 2 : The comparison of the deep learning methods

Method	Description	Advantages	Disadvantages
Deep Neural Network [22]	Deep neural Network is a technique using neural network which uses some layers of nodes between the input and output. These layers are used to identify the features in a videos and processes into some stages. The deep neural network is motivated by the human brain. Its function is similar to human brains. This deep learning method is used for classification of an object and pattern analysis. This method is focused to recognize something which can be a human face, voice, object and etc. The application of this system is voice recognition system, human recognition system and face recognition system. So the researchers choose this technique to recognize human in video surveillance system.	• Has a larger set of functions • Trainable from beginning to the end • It has the ability to learn complex function	• Needs large amount of data • Expensive to train • Learning deep learning method is hard

Method	Description	Advantages	Disadvantages
Haar-Like Features [22]	Haar-Like features is a characteristic that is used for digital image. It is mainly used for object recognition system. It works well with image intensities which is in RGB pixel values. This method makes all the calculations to extract the features from an image or video. Then, it also has the ability to count the features of any size in a constant time. Some of the researchers use this method because it can be easily trained to extract specific features from an object.	• It has the highest rate of calculation speed • Any size of image can be used	• Does not allow real time detection • Sensitive in different poses
Viola Jones [23]	Viola Jones is a real time object detection framework that was proposed by Paul Viola and Michael Jones. It also can be taught to detect and recognize either human or object. It has a very high detection rate and the real time detection speed is 2 frames per second. This method has four stages to perform in one task which is first Haar feature selection, second creating an integral image, third Adaboost training and the last stage is cascading the classifier. This Viola Jones in object tracking has the ability to recognize static and moving object.	• Can classify moving objects • Works on real time system to recognize	• Difficult to find the features • Corrupt when occlusion happen

Method	Description	Advantages	Disadvantages
Convolutional Neural Network (CNN) [23]	Convolutional Neural Network (CNN) is a classifier and it has the ability to perform in visual recognition. This CNN has two layers which is convolutional layer and pooling layer. Convolutional layer is used to filter the signals. The pooling layer is used to reduce the amount of spacial size. The CNN architecture contains input layer, convolutional layer, pooling layer and output layer. This layers act as a learnable features extractor. The more the layers, the accurate the system is. To train the CNN method, there are three steps which is data preparation, model definition and solver definition. First of all, we have to prepare some data for this method such as the arrangement of the layers, number of neurons, weight of the system and etc. Many researchers focus in this method because it is able to recognize human if the system has a strong dataset.	• It is an effective tool used to recognize object • Simple to implement • Learnable to classify data	• High computational cost • Need a lot of data

2.2.2.7 Deep Learning Frameworks Comparison [24]

Table 2. 3: The comparison of the Framework (API)

Framework	Description	Advantage	Disadvantage
Tensorflow	Tensorflow is a library and open source software to make the numerical calculation by using data graph	• It can easily handle the videos, images and audios • Large amount of codes and models are available	• The neural network lacking • Needs large amount of memory
Theano	Theano is one of the python library. It uses a computational graph to produce an output.	• The optimization is fast and stable • It can automatically differentiate • Saves a lot of codes without damaging flexibility	• Issues when compiling • A lot of error messages pop up
Keras	Keras is high level library of the neural network. It can be used easily in Python.	• Easily create new modules • Designed for human being and it is user-friendly • Easy to use in real time models	• It is very less flexible • Less projects • Multiple usage of GPU
Caffe	Caffe is mainly used in deep learning to apply in vision, multimedia and speech.	• Easy in coding • Fastest library in CPU • Easy to combine many libraries	• Needs large amount of memory

Table 2. 4: The comparison of the Framework (API)

Framework	Language(Coding)	Community	Documentation	Performance
Tensorflow	• Microsoft visual(C++), Matlab, Java, python • Multi GPU, Windows, Android	• It is large and also growing	• Low	• High
Theano	• Python • Multi GPU	• Low community	• High	• Medium
Keras	• Python • Multi GPU	• Growing	• High	• Medium
Caffe	• Microsoft visual(C++), Matlab, python • Multi GPU, Windows	• Largest community	• Low	• Very high

CHAPTER 3

METHODOLOGY

In this section, the methods and techniques that are implemented in accomplishing this project is discussed. Hence, the method and tools that are to be used must be properly planned and chosen to ensure that the project will be successful within the time required and fulfil its objectives that have been listed before. The previous studies enabled to choose the most suitable technique and understand the project's views. Development of human recognition system must have the knowledge about Machine Vision and Artificial Intelligence. This system needs software and libraries to design the human pedestrian detection system which can analyse and classify the videos that are being recorded. This system will face some problems when detecting human pedestrians such as distance of the camera, light intensity and etc. So the system will be analysed using different light intensities, different environments and occluded human that passes by the camera.

3.1 Flow chart

In this part, discussion about the steps and methods included in this project is presented in the form of a flow chart. There are two types of flow chart, which is an overall planning flow chart and a system planning flow chart. The overall planning flow chart will briefly show the process of the whole final year project system. The system planning flow chart will explain about the steps of this system.

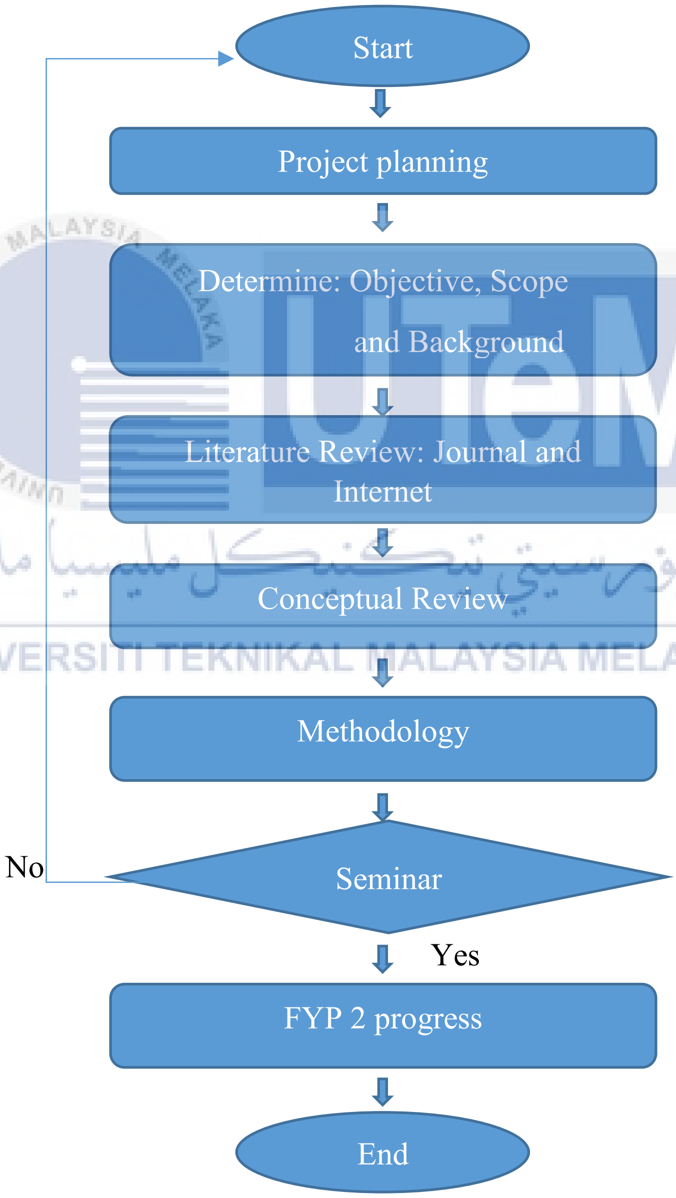


Figure 3. 1 : The Overall flow chart (FYP 2)

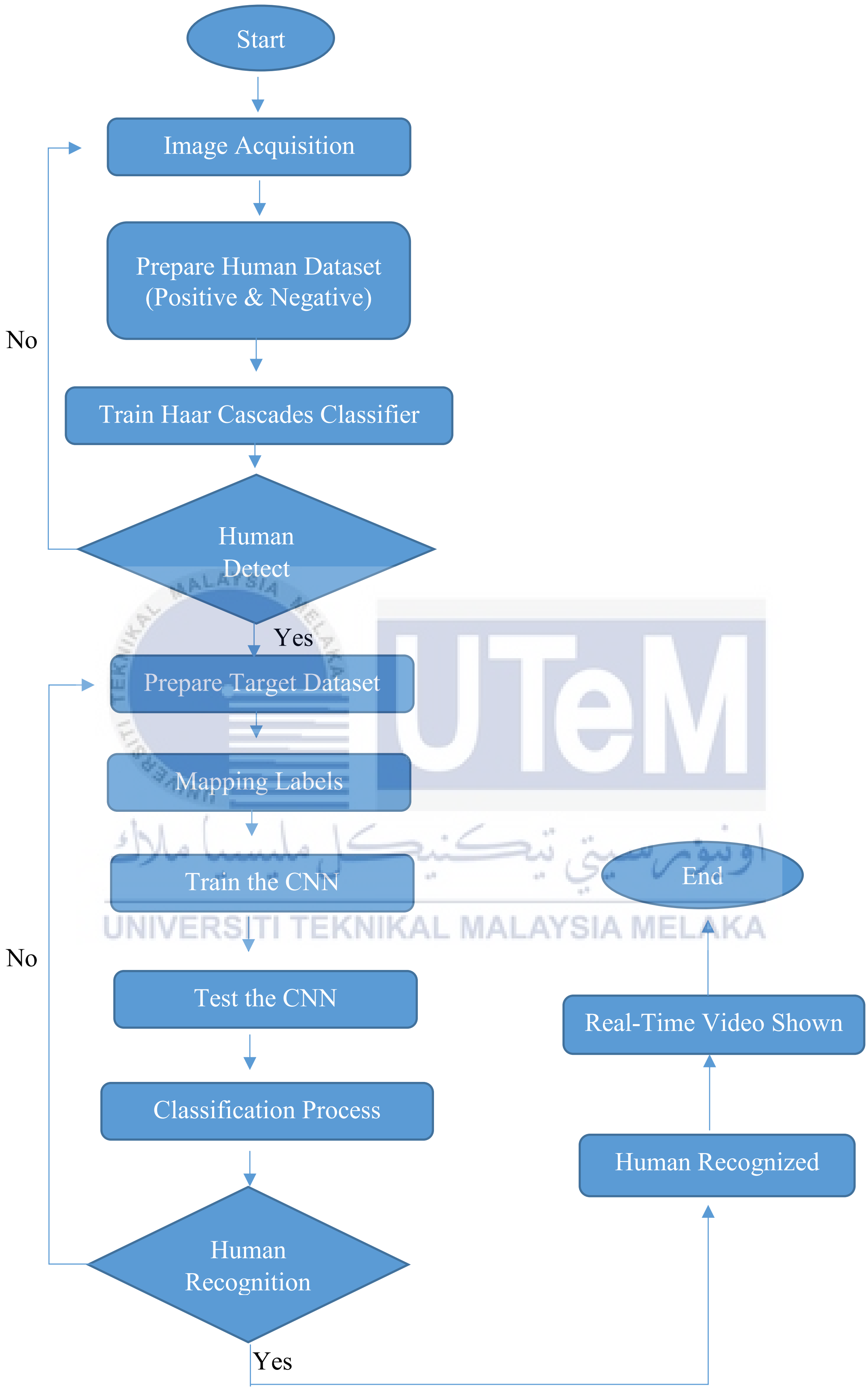


Figure 3. 1: The system flow chart

3.2 Hardware part

This section will explain about the hardware that will be used in this system. In the previous researches, some of the researchers used Arduino, Personal computers and Raspberry pi as the controller. For this project, fast and user friendly controller is required. Because of that, the Personal Computer has been chosen as a controller because it is faster compared to the other controllers. Webcam camera with (5MP) will be used for the vision of this system. The comparison of the controller is as follow:

Table 3. 1: Comparison of the controller [10]

Controller	Description	Speed	Space
Personal Computer	• Can perform any kind of tasks given • Easy to use and is the fastest controller	Very High	Based on the external hard disk
Arduino	• Has the ability to store and run one program at a time • Difficult to use in image processing techniques	Moderate	32kb
Raspberry Pi	• Has the ability to store and run multiple programs at a time • It is slow in training or installing libraries	Moderate	Based on the SD Card size

3.2.1 Personal Computer / Laptop

Personal computer is a controller that is widely used nowadays by everyone to make any tasks easier. It is very user friendly and is the fastest controller to perform any kind of tasks associated. A strong and fastest controller is required for this system to perform multiple tasks at the same time. This system also have to train thousands of data to perform the human recognition task. Therefore, high data space is required for this system. From the comparison, it can be surely said that the personal computer is very suitable for this system. As an advantage, RAM can be added in the computer to speed up the system.

3.2.2 Webcam

Webcam is a video camera that will act as a sensor (vision) or eyes mainly for the computer. It will record and capture real time video or streaming video and image. It is very easy to setup a webcam and very cheap in market price. Logitech (5MP) webcam camera is chosen for this system. This camera has the ability to track the motion of objects and has 720P resolution to capture the videos. Then, it is also very fast in tracking an object in a video stream and has a good lighting even when it is in the situation of a dim light. The videos that are captured by this camera are very clear and easy to perform in real time situations.

3.3 Software Part

There are three types of software that can be used in this human pedestrian recognition system which is Matlab, Microsoft Visual Basic and Python. All the coding of this system will be performed by these software. Table 3.2 shows the comparison of these software.

Table 3. 2: The comparison of the software [25]

Software	Library	Documentation	Community	Cost	Learning
Matlab	Powerful matrix library	Great documentation	Low	High	Hard to learn
Microsoft Visual Basic	Huge optimized library	Low	Big	Open source	Difficult for beginner
Python	Easy to use and huge number of library	Weak	High	Open source	Suitable for beginner

From the comparison made, Python has been chosen as a software that will be used in this whole project. It is the easiest language that can be learnt compared to other software. It also has the highest community where any helps can be obtained through online. Python is an open source so it has no payment or licence to use it. Developer can easily apply any library needed in this software. Python will be the main software for this system. It also will be used for the analysis of this system. Two types of Python versions are available that is 2.7 and 3.3. For this system, Python 3.3 is the most suitable version and it has the latest features.

3.3.1 Framework (API)

Table 3. 3: The comparison of Framework [24]

Framework	Language (Coding)	Community	Documentation	Performance
Tensorflow	• Microsoft visual (C++), Matlab, Java, Python • Multi GPU, Windows, Android	• It is large and also growing	• Low	• High
Theano	• Python • Multi GPU	• Low community	• High	• Medium
Keras	• Python • Multi GPU	• Growing	• High	• Medium
Caffe	• Microsoft visual (C++), Matlab, Python • Multi GPU, Windows	• Largest community	• Low	• Very high

Keras is the library that will be used in this whole system to create and train the datasets. Keras is a high level network API, which is written in Python and capable of running on Tensorflow and Theano. Keras is user friendly and it is designed for human beings not for the machines. It also has easy extensibility. By using Keras, it can easily create a model and at the same time can be add new models for this system. Before installing Keras, installation of the backend engines, which is Tensorflow backend and Theano backend is needed. Backend engine means an algorithm that make all the calculations. Keras is also a most popular and fast growing deep learning framework. Nowadays, Keras framework is used to tackle real world problem ranging from computer vision to language processing. So, it is very suitable for this human recognition system.

3.4 System Construction

This section will discuss about the system construction and the flow of the system. It will explain about the method and technique used to complete this project.

3.4.1 Image Acquisition

The image acquisition is the first stage of any vision system. This method is used to detect the moving objects by performing a pixel to pixel comparison. If there is no moving objects, the pixel values will be the same for each of the frame. There is a software for this image acquisition method which is Image Acquisition Toolbox. There are some steps to setup the Image Acquisition software which at first, install and configure the image Acquisition device. Second, retrieve the hardware information. Then, create a video input object and preview the video stream. After that, configure the object properties and acquire the image data. Then, the last step is clearing the memory after the image Acquisition object process is done. The image acquisition will capture the videos of the moving objects and convert it into a measurable energy. The energy is light or can be said as electromagnetic waves. There are three parts in this image acquisition method which is energy reflecting process, optical system and energy measurement. The errors of lighting can be reduced when Image Acquisition happens. So for this project, need to do the image Acquisition as the first step. The equations for the Image Acquisition are as below:

$\gamma = c/f$, E=h.f E= (h.c)/ γ (3.1)

γ= Wavelength

c= Speed of the light

E=Energy

f=Frequency

h= Planck's constant

3.4.2 Haar Cascades

Haar cascades are cascade classifiers based on Haar Features. Cascade classifier is concatenation of a set of weak classifiers used to create a strong classifier. Weak classifier are the classifiers with limited performance. It does not have the ability to classify everything correctly. If a system gives a simple problem, it might perform but in acceptable level. Strong classifiers are really good at classifying data accurately. Another part of Haar Cascades is Haar features. This Haar features are simple summation of rectangles and differences of those areas across the image. This method is very accurate in detection of objects. In 2001 Paul Viola and Jones published a paper which states about a fast and very accurate method in object detection. They use Boosted Cascade of simple classifiers. It is used to build strong classifiers that can perform really well in detection. They used simple classifier to maximize the problems when building single classifier and to make the classifier perform with high precision. The single step classifiers tend to become complex and computationally intensive. These Haar cascade perform very well in detection because it uses simple classifiers which that is not complex in detection. Rectangular features can be calculated by the equation below:

$$ii(x,y) = \sum_{x' \le x, y' \le y} i(x',y'), \qquad (3.1)$$

$ii(x,y)$ = integral image

$i(x',y')$= Original image

Calculation:

$$s(x,y) = s(x,y-1) + i(x,y) \qquad (3.2)$$

$$ii(x,y) = ii(x-1,y) + s(x,y) \qquad (3.3)$$

$s(x,y)$ = cumulative row sum

$ii(x-1,y)$ = 0, the integral image can be computed in one pass over the original image

3.4.3 Deep Learning

Deep learning is the main technique that will be used in this human recognition system. It is a class of neural network models. A model has an input layer, an output layer and a number of hidden layers. These layers are made up with neurons. The behaviour of the neurons are inspired by the human brains neurons. Neurons are the part of any deep learning methods. Figure below explains each of the flow of implementation of the deep learning in this human pedestrian recognition system:

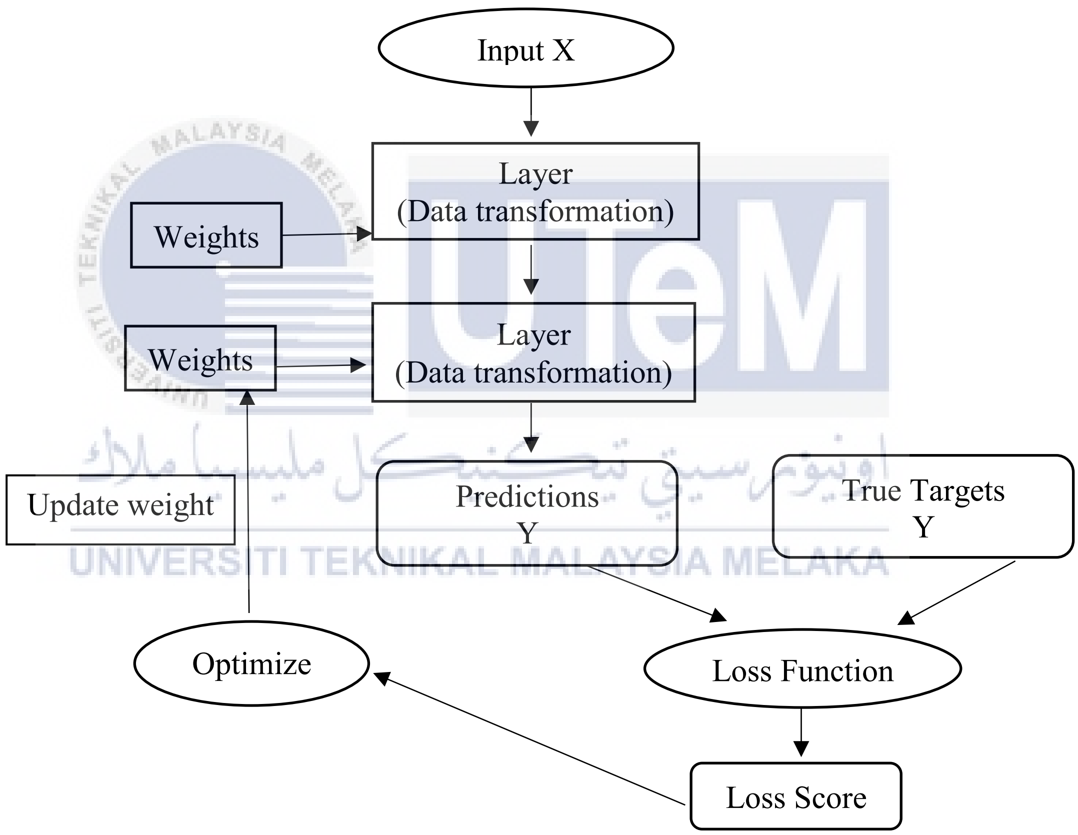


Figure 3. 2: The flow of implementation of Deep Leaning in this recognition system

Input X are the input of the system, which is the human images. The input X will convert the inputs to the targets such as (label "sasi"). Layer is called as simple data transformation. All the data of the input will be filtered and learned by exposure. The specifications that are filtered by the layers will be stored in layer's weights, which are in a number form. Weight is the parameter of the layer. The learning in the deep learning means finding a fixed value for the weights of all layers in a network. So that the network can correctly declare the inputs to appropriate targets. Then, the Prediction Y will predict the inputs. It will take the highest probability of the targets to predict. Besides, Loss Function is also called as objective function. It will measure how far the output is from what is expected. It will take the prediction of the network and what the system wanted the network to output (True Target) and computes a distance score, tells that how well the network has done. If the rate of loss function is not as expected, a feedback signal will adjust the value of the weight a little to reduce the loss score. This adjustment is done by the Optimizer by using the Backpropagation algorithm. Normally if the output is not as what expected means the loss score might be high. The optimizer will decrease the loss function by adjusting the weight automatically. It also finds a very suitable weight values. This training loop will be repeated for several times to reduce the yield weight values that minimize the loss function. The minimal loss function is the one close to the expected target.

3.4.3.1 The neural network

The neural network is the full architecture by the layers. There are three type of layers which is the input layer, hidden layer and output layers. The convolutional neural network (CNN) is the one widely used in all computer vision tasks. Each layers work as filter for the input image. The last layer will be used to predict the probability of the image. The Convolutional Neural networks will be used for this human recognition system to classify the human. The more the images in the dataset, the more accurate the system.

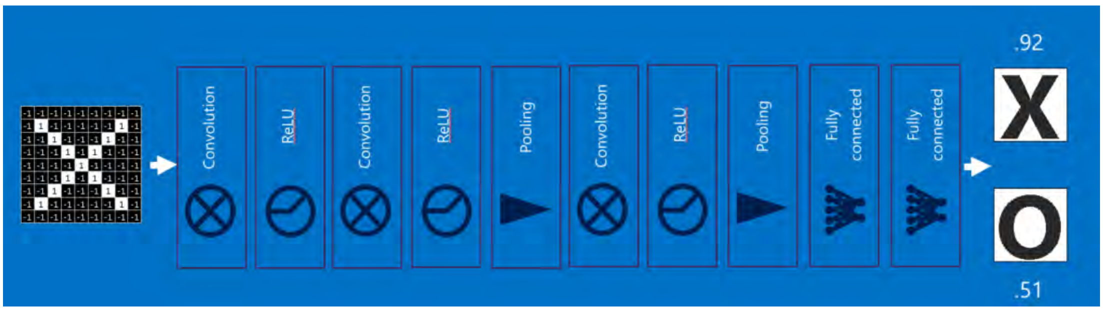


Figure 3. 3: The Layers in Convolutional Neural Network

3.4.3.2 Neuron Model [26]

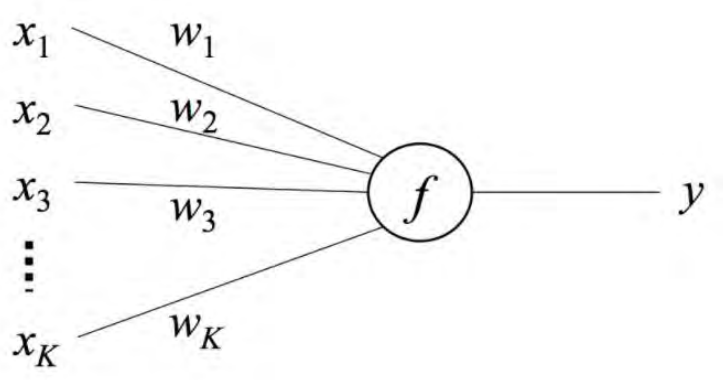


Figure 3. 4: The Neuron Model

X_1-X_k: Input vector

W_1-W_k: weight vector

F: Nonlinear function

Y: scalar output (prediction)

f: takes a weighted sum of input and return y. Input vector always equals to 1. f is called as link function between input and output. Link Function is called as logistic function.

$$y = f(\sum_{i=0}^{K} w_i x_i) = f(\mathbf{w}^{\mathbf{T}}\mathbf{x}) \quad (3.4)$$

$$\text{Logic function} = f(u) = \frac{1}{1+e^{u}} \quad (3.5)$$

$$\text{Single neuron model} = y = \frac{1}{1+e^{\mathbf{w}^{\mathbf{T}}\mathbf{x}}} \quad (3.6)$$

3.4.3.3 The Neural Network Architecture

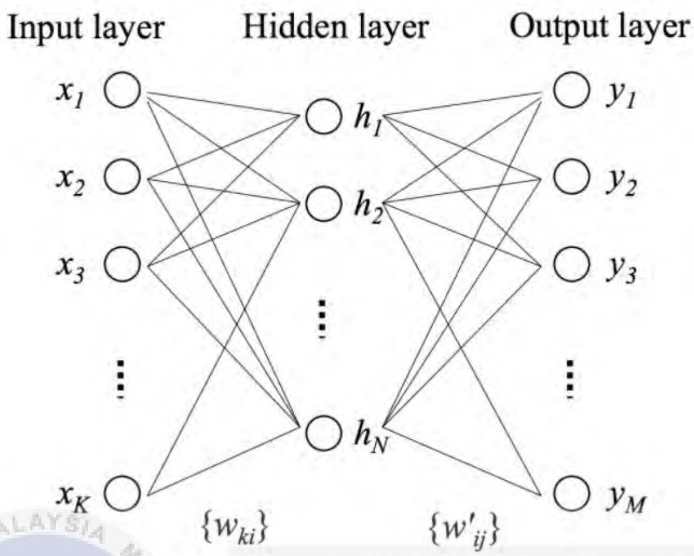


Figure 3. 5: The Neural Network of Deep Learning[26]

X=(X1-Xk) = Input Vector

h= (h1-hN) = Vector of hidden neurons

y= (y1-yM) =Output vector

w_{ki}= weight refers with the connection between the k^{th} input element and the i^{th} Hidden layer

w'_{ij}= Weight between i^{th} hidden layer and j^{th} output layer

The input layer consist of the input vector, the hidden layers consist on vector of neurons and the output layer consist of the output vector. Every elements of the input layers are connected with the hidden layers and every element of the hidden layers are connected to the output layers.

3.4.3.4 The derivation of neural network equation [26]

Start from Chain Rule:

$$\frac{\partial E}{\partial w'_{ij}} = \frac{\partial E}{\partial y_j} \cdot \frac{\partial y_j}{\partial u'_j} \cdot \frac{\partial u'_j}{\partial w'_{ij}} \tag{3.7}$$

The derivative E with respect to y_j

$$\frac{\partial E}{\partial y_j} = y_j - t_j \tag{3.8}$$

The logistic Function f with respect to the input u is $f(u)(1 - f(u))$

Logistic Function= $\frac{df(u)}{du} = f(u)(1 - f(u)) = f(u)f(-u)$ (3.9)

After apply this get:

$$\frac{\partial y_j}{\partial u'_j} = y_j(1 - y_j) \tag{3.10}$$

Computer the derivative of $u'_j = \sum_{i=1}^{N} w'_{ij} h_i$ with respect to w'_{ij}

$$y_j = f(u'_j)$$

$$\frac{\partial E}{\partial w'_{ij}} = (y_j - t_j) \cdot y_j(1 - y_j) \cdot h_i \tag{3.11}$$

With the gradient construct the update equation of w'_{ij}

$$w'^{new}_{ij} = w'^{old}_{ij} - \eta \cdot (y_j - t_j) \cdot y_j(1 - y_j) \cdot h_i \tag{3.12}$$

By using the chain rule, the full gradient is:

$$\frac{\partial E}{\partial w_{ki}} = \sum_{j=1}^{M} \left(\frac{\partial E}{\partial y_j} \cdot \frac{\partial y_j}{\partial u'_j} \cdot \frac{\partial u'_j}{\partial h_i}\right) \cdot \frac{\partial h_i}{\partial u_i} \cdot \frac{\partial u_i}{\partial w_{ki}} \tag{3.13}$$

Computed $\frac{\partial E}{\partial y_j}$ and $\frac{\partial y_j}{\partial u'_j}$

$$\frac{\partial E}{\partial y_j} \cdot \frac{\partial y_j}{\partial u'_j} = (y_j - t_j) \cdot y_j(1 - y_j) \tag{3.14}$$

Compute the remaining derivatives $\frac{\partial u'_j}{\partial h_i}$, $\frac{\partial h_i}{\partial u_i}$ and $\frac{\partial u_i}{\partial w_{ki}}$

$$\frac{\partial u'_j}{\partial h_i} = \frac{\partial \sum_{i=1}^{N} \quad w'_{ij} h_i}{\partial h_i} = w'_{ij} \tag{3.15}$$

$$\frac{\partial h_i}{\partial u_i} = h_i(1 - h_i) \tag{3.16}$$

$$\frac{\partial u_i}{\partial w_{ki}} = \frac{\partial \sum_{k=1}^{K} \quad w_{ki} x_k}{\partial w_{ki}} = x_k \tag{3.17}$$

Making Substitution at the gradient equation

$$\frac{\partial E}{\partial w_{ki}} = \sum_{j=1}^{M} \quad [(y_j - t_j) \cdot y_j(1 - y_j) \cdot w'_{ij}] \cdot h_i(1 - h_i) \cdot x_k \tag{3.18}$$

Update the Equation

$$w_{ki}^{new} = w_{ki}^{old} - \eta \cdot \sum_{j=1}^{M} \quad [(y_j - t_j) \cdot y_j(1 - y_j) \cdot w'_{ij}] \cdot h_i(1 - h_i) \cdot x_k \tag{3.19}$$

This derivative is also called as backpropagation because we begin with the final output error for output neuron and this error gets propagated backwards throughout the networks in order to update the weight. The backpropagation algorithm is used to modify the error of the system automatically by reducing or increasing the weight range. It also find right values for the weight.

3.5 Experimental Setup

This section will introduce all the experiments that has been conducted in this system and the parameters that are included in this experiments. In this system, three experiments have been carried out. This section will explain the objectives and procedures of those experiment. Table 3.4 shows the list of the experiment:

Table 3. 4: The list of the experiment

NO	EXPERIMENT
1	Analysis of light intensity when the camera recognize the occluded human
2	Analysis of the distance between object and camera when recognize the occluded human
3	Analysis of accuracy and performance when the camera recognize the occluded human

3.5.1 Constant values and Equations

3.5.1.1 Constant values

1. **Distance from ground to camera = 2 meter**

The 2meter distance has been chosen because it is the most suitable distance to place the camera and from this distance the camera able to capture the image of human fully.

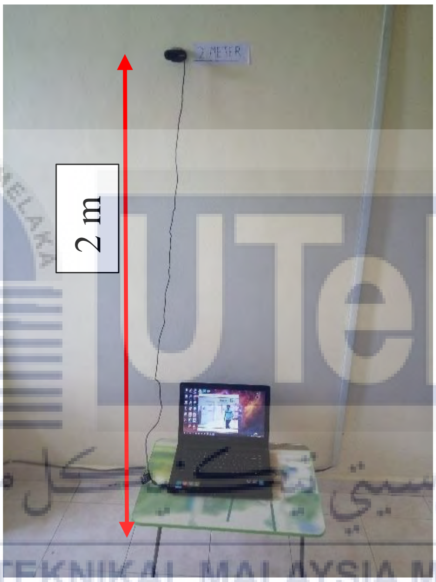


Figure 3. 6: Experimental setup of camera

2. **Frame rate of the video(fps)**

Frame rate is the frequency rate which at consecutive images called frames appear on display. This system uses constant frame rate of 60 fps for the video streams. By using the OpenCV library, the frame rate is set to a specific value. This system had been set to 60fps to make the video stream faster without any interruption. 60fps is smoother compared to other frame rates. The camera used in this system has ability only to capture the video in 60 frame per second. So, the 60 fps the most suitable for this system.

3.5.1.2 Experimental equations

True Positive Rate$(TPR) = TP/(TP + FN)$ (3.20)

True Negative Rate $(TNR) = TN/(FP + TN)$ (3.21)

False Positive Rate $(FPR) = FP/(FP + TN)$ (3.22)

False Negative Rate $(FNR) = 1 - TPR$ (3.23)

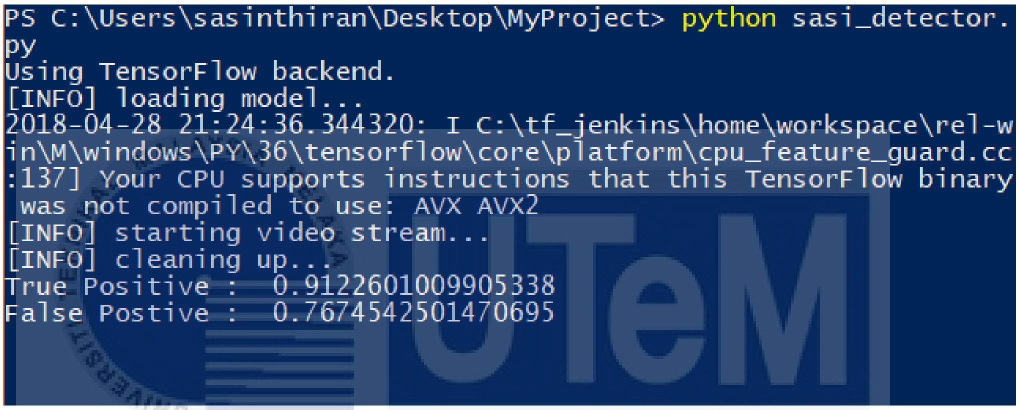


Figure 3. 7: Console showed the experimental values

Miss Rate = $FP/(TP + FP)$ (3.24)

Total Rate of Error = Total Miss Rate / Total Number of test (3.25)

Accuracy= 1- Total Rate of Error (3.26)

Performance= 1- [(Actual accuracy – Desired Accuracy) / (Actual Accuracy)] (3.27)

Mean = $\bar{x} = (\Sigma x_i) / n$ (3.28)

3.5.2 Experiments and Procedures

3.5.2.1 Experiment 1: Analysis of the distance between object and camera when recognize the occluded human (indoor)

Objective:

1. To find out the accuracy and performance of the system when occlusion happens with different distance between camera and pedestrian
2. To determine the area that this system covered when the occlusion happens
3. To find out the rate of error of the system when different distance

Procedures:

1. The camera was setup 2m from the ground
2. The light intensity was taken based on the indoor (520Lux)
3. Once the system start, it will show the real time video and capture the pedestrian that pass by it
4. The video was analysed and tabulated
5. Step 3 and step 4 repeated 20 times with different distance which is 2m, 3m and 4m

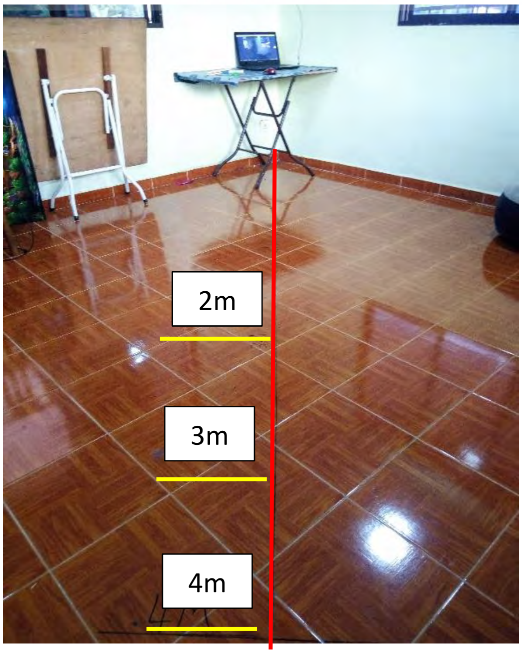


Figure 3. 8: The different level of distance used in this experiment

3.5.2.2 Experiment 2: Analysis of light intensity when the camera recognize the occluded human (indoor)

Objective:

1. To find out the accuracy and performance of the system when occlusion happens with different lighting
2. To determine the best light intensity for the human recognition system
3. To find out the rate of error of the system when in different lighting

Procedure:

1. The camera was setup 2m from the ground
2. The distance between human and camera is fixed to 4m
3. Once the system start, it will show the real time video and capture the pedestrian that pass by it
4. The video was analysed and tabulated
5. Step 3 and step 4 repeated 20 times with different light intensity

Table 3. 5: List of light intensity used in this system

Light Intensity(Lux)	Level
320 Lux	Dim Light
520 Lux	Moderate Light
750 Lux	Very Bright

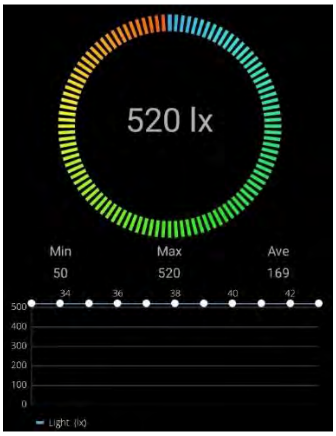


Figure 3. 9: Moderate light (indoor)

3.5.2.3 Experiment 3: Analysis of accuracy and performance when the camera recognize the occluded human (indoor)

Objective:

1. To determine the accuracy of recognition when the occlusion happens
2. To determine the performance of this system when the occlusion happens
3. To find out the overall rate of error when the occlusion happens

Procedure:

1. The camera was setup 2m from ground
2. The light intensity was taken based on the indoor (520 lux)
3. The distance from camera to the pedestrian is 2m
4. Once the system start, it will show the real time video and capture the pedestrian that pass by it
5. The video was analysed and tabulated
6. Step 3 and step 4 repeated 20 times with same distance, different pedestrian and environment.

Table 3. 6: Table of Level of accuracy

Percentage of Accuracy (%)	Level
99.5-95	Independent Level (Very Accurate)
94-90	Instructional Level (Medium)
89-50	Frustration (Low)

CHAPTER 4

RESULT

This chapter will explain about the results and data obtained from the experiments that have been conducted before this. All the experiments are conducted in indoor environment because the scope of this project is for indoor environment. In this environment, several parameters are tested to find out the accuracy of the system. From the experiment, several data has been collected and tabulated into tables and graphs. The experiments are very important to measure a system's performance and accuracy. All data obtained from the experiment is analysed by a graph.

4.1 Testing System

The system need to be tested once the design of the system is complete. Testing is very important for a project because it enables to see how well the system is working and also get an outcome. The outcome can prove whether the system is able to work properly or not. For this project, two tests have been conducted to find out the outcome of this system. These tests are not a part of the experiment. It's just a trial session for the system to check whether it works in detecting human or not. One of the tests that has been conducted is testing the detection of pedestrian that passes by the camera. Then, the second one is testing the system with recognition of pedestrian. From the tests, it enables to adjust the programs, algorithms and techniques if the result is unsatisfactory. Once the test shows a great outcome, then we may start to do the experiment. These tests should follow our objective which is based on the human detection and recognition. Every test will be conducted in indoor environment because the analysis will be performed in indoor environment. Table 4.1 shows the list of tests.

Table 4. 1: The list of the test

No	Test
1	Detection of pedestrian that passes by the camera
2	Recognition of pedestrian that passes by the camera

4.1.1 Test 1: Detection of pedestrian that passes by the camera

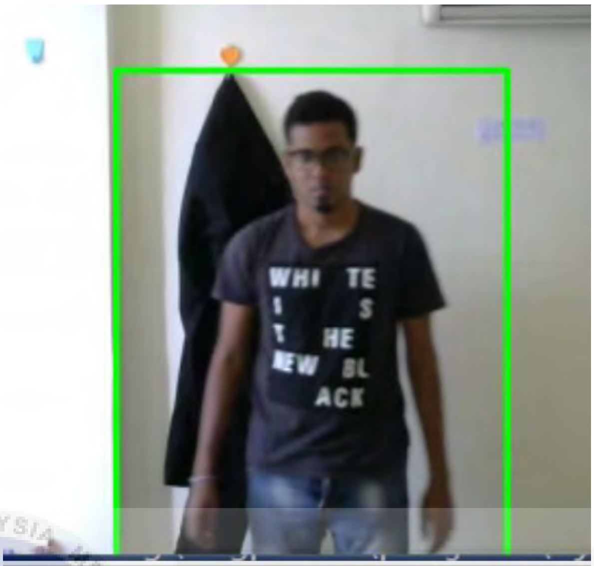


Figure 4. 1: The detection of pedestrian

Figure 4.1 shows one of the image of the pedestrian taken from the video of detection system while testing. Figure shows one person appearing with a bounding box when passing by the camera. The bounding box shows that the system is able to detect or track the body of the person. However, even though the detection is quite accurate, there are also some false positive caused by the shadows of the humans. It shows that light intensity is very important for this system. This system focuses on real time detection. So, the video processing is more difficult compared to the image processing. It needs a quick response to perform the system.

4.1.2 Test 2: Recognition of pedestrian that passes by the camera

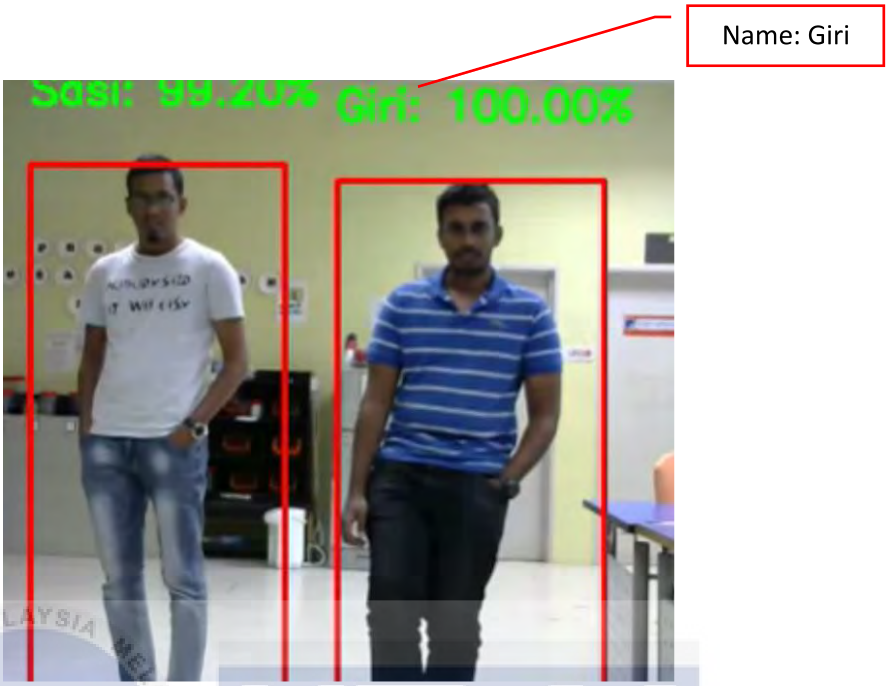


Figure 4. 2: The recognition of pedestrian

Figure 4.2 shows the outcome of recognition test. It clearly shows that this system is able to track and recognize humans. Name of the human is displayed on top of the bounding box if the system recognizes the human. The percentage is the amount of possibility that matches the recognition of human. However, even though the recognition is quite accurate, there are also some false positive caused by the light intensity. The light must be balanced, if not the recognition task will be difficult for the computers to conduct. So, this system must be trained by using more data to encounter this problem. Otherwise, this system perform very well in recognition. From the test, this system uses more data of the pedestrians to make it more accurate.

4.2 Classification of Detection for Occlusion

In this system, the images will be classified into four types of occluded detection which is True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Every calculation will be made by using these types of detection. From the system, the values of True Positive (TP) and False Positive (FP) can be obtained in percentage. Then, the values of the True Negative (TN) and False Negative (FN) need to be calculated by the equation of detection which is (FP+FN) = 100% - (TP + TN). TP is the successful detection where the bounding box will appear correctly to each and every person that passes by this system. If the bounding box is able to detect each person that passes by the camera of the system it means it’s the TP detection.

Then, the second one is TN. It means that the system detects both person in one bounding box which means it detects the occluded person and the other person in one frame and also shows that only one person appears in the camera. The next one is FP which means it detects each and every person passing by the camera and also some extra unknown things such as shadow. This is where the bounding box appears around the real person and also at empty spaces where it detects the extra unknown things. Lastly, FN means that the system is not able to track anything. It shows no bounding boxes in the system. The figure below shows the example of the type of detections.

Figure 4. 3: TP type of detection

Figure 4. 4: TN type of detection

Figure 4. 5: FP type of detection

Figure 4. 6: FN type of detection

4.3 Data Observation

This subtopic will show the results and data of the experiments that have been conducted which is accuracy test, distance test and light intensity test. The experiments follow the procedure that has been planned in chapter 3. From the experiments, the data is recorded, graph is plotted and analysis is made.

4.3.1 Analysis of the Distance between object and camera when recognize the occluded human

4.3.1.1 Distance test of 2meter from camera to pedestrian

Table 4. 2: Result of distance test (2meter)

Person	Place	TP	TN	FN	FP	Miss Rate/Error
1	Library	0.9252	0	0.0748	0.0408	0.0422
2		0.9265	0	0.0735	0.0524	0.0535
3		0.9185	0	0.0815	0.0328	0.0345
4		0.9365	0	0.0635	0.0425	0.0434
5		0.9375	0	0.0625	0.0385	0.0394
6	Class room	0.9025	0	0.0975	0.0515	0.0540
7		0.9255	0	0.0745	0.0423	0.0437
8		0.9168	0	0.0832	0.0125	0.0135
9		0.9321	0	0.0679	0.0265	0.0276
10		0.9420	0	0.0580	0.0285	0.0294
11	House	0.9388	0	0.0612	0.0375	0.0384
12		0.9275	0	0.0725	0.0565	0.0574
13		0.9395	0	0.0605	0.0482	0.0488
14		0.9285	0	0.0715	0.0310	0.0323
15		0.9490	0	0.0510	0.0365	0.0370
16	Canteen	0.9186	0	0.0814	0.0425	0.0442
17		0.9244	0	0.0756	0.0512	0.0525
18		0.9528	0	0.0472	0.0379	0.0383
19		0.9355	0	0.0645	0.0452	0.0461
20		0.9468	0	0.0532	0.0396	0.0401
Total		18.6245	0	1.3755	0.7944	0.8163

<u>Constant</u>

- Distance from ground to camera : 2m
- Frame Rate: 60fps
- Distance from camera to pedestrian : 2m
- Light Intensity: 520 LUX

<u>Example of miss rate calculation</u>

Miss Rate = $FP/(TP + FP)$

Miss Rate= 0.0408/ (0.9252+0.0408) = 0.0422

Total miss rate = 0.8163

Total Rate of Error = Total Miss Rate / Total Number of test

Total Rate of Error= 0.8163/20= 0.0408

Accuracy= 1- Total Rate of Error

Accuracy= (1- 0.0408) X (100) = 95.92%

Actual Accuracy=0.9735

Desired Accuracy=0.9592

Performance= 1- [(Actual accuracy – Desired Accuracy) / (Actual Accuracy)]

Performance= 1-[(0.9735-0.9592)/ (0.9735)]

Performance=0.9853x100%= 98.53%

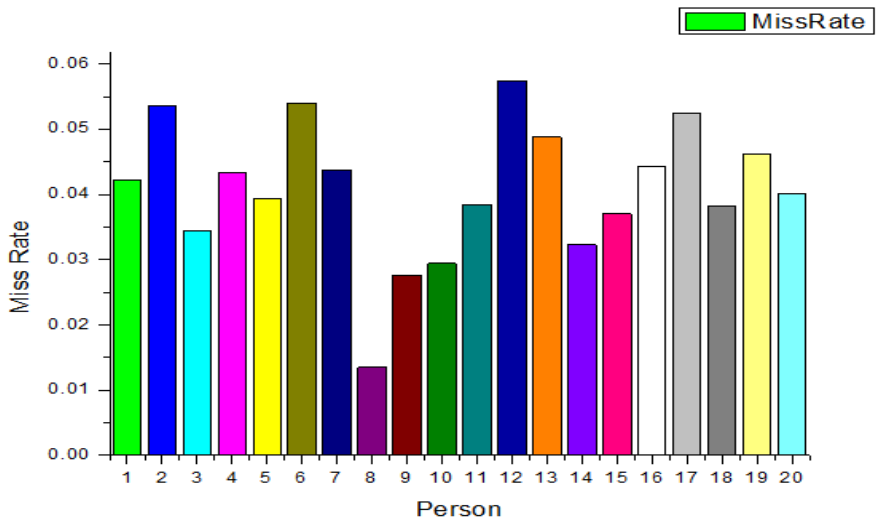


Figure 4. 7: Histogram of the miss rate when the distance is 2m

4.3.1.2 Distance test of 3meter from camera to pedestrian

Table 4. 3: Result of distance test (3meter)

Person	Place	TP	TN	FN	FP	Miss Rate/Error
1	Library	0.9056	0	0.0944	0.0988	0.0983
2		0.8871	0	0.1129	0.1230	0.1217
3		0.8752	0	0.1248	0.1150	0.1161
4		0.9016	0	0.0984	0.1050	0.1043
5		0.8625	0	0.1375	0.1282	0.1294
6	Class Room	0.8165	0	0.1835	0.1475	0.1530
7		0.8511	0	0.1489	0.1492	0.1492
8		0.8728	0	0.1272	0.1289	0.1287
9		0.8352	0	0.1648	0.1755	0.1736
10		0.8765	0	0.1235	0.1255	0.1252
11	House	0.8977	0	0.1023	0.1268	0.1227
12		0.8465	0	0.1535	0.1678	0.1654
13		0.9025	0	0.0975	0.1205	0.1019
14		0.9005	0	0.0995	0.0886	0.0896
15		0.8556	0	0.1444	0.1235	0.1261
16	Canteen	0.8775	0	0.1225	0.1465	0.1431
17		0.8588	0	0.1412	0.1568	0.1544
18		0.8952	0	0.1048	0.0995	0.1001
19		0.8735	0	0.1265	0.1285	0.1282
20		0.8967	0	0.1033	0.1165	0.1097
Total		17.4886	0	2.5114	2.5523	2.6838

<u>Constant</u>

- Distance from ground to camera : 2m
- Frame Rate: 60fps
- Distance from camera to pedestrian : 3m
- Light Intensity: 520 LUX

<u>Example of miss rate calculation</u>

Miss Rate = $FP/(TP + FP)$

Miss Rate= 0.0988/ (0.9056+0.0988) = 0.0984

Total miss rate = 2.6838

Total Rate of Error = Total Miss Rate / Total Number of test

Total Rate of Error= 2.6838/20= 0.1342

Accuracy= 1- Total Rate of Error

Accuracy= (1- 0.1342) X (100) = 86.58%

Actual Accuracy=0.9735

Desired Accuracy=0.8658

Performance= 1- [(Actual accuracy – Desired Accuracy) / (Actual Accuracy)]

Performance= 1-[(0.9735-0.8658)/ (0.9735)]

Performance=0.8894x100%= 88.94%

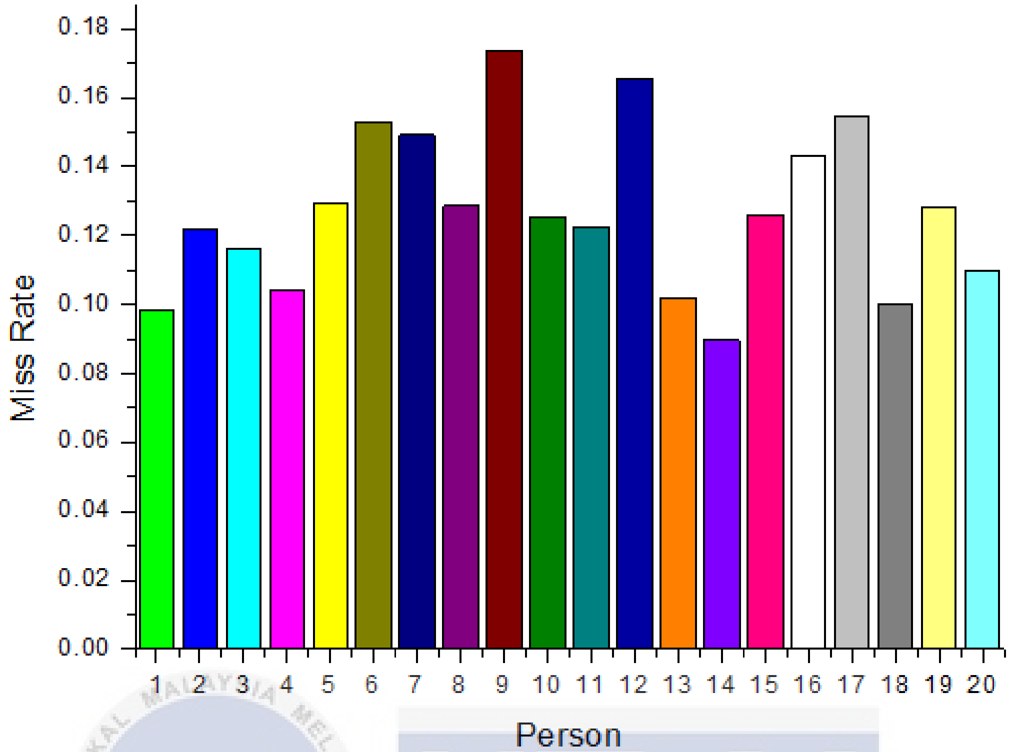


Figure 4. 8: Histogram of the miss rate when the distance is 3m

4.3.1.3 Distance test of 4meter from camera to pedestrian

Table 4. 4: Result of distance test (4meter)

Person	Place	TP	TN	FN	FP	Miss Rate/Error
1	Library	0.6252	0	0.3748	0.4885	0.4386
2		0.6116	0	0.3884	0.5123	0.4558
3		0.5843	0	0.4157	0.5245	0.4730
4		0.5765	0	0.4235	0.4899	0.4594
5		0.5024	0	0.4976	0.5018	0.4997
6	Class Room	0.5118	0	0.4882	0.5124	0.5003
7		0.4785	0	0.5215	0.5524	0.5358
8		0.5321	0	0.4679	0.5043	0.4866
9		0.4644	0	0.5356	0.5684	0.5503
10		0.4585	0	0.5415	0.5898	0.5626
11	House	0.5101	0	0.4899	0.6421	0.5573
12		0.5432	0	0.4568	0.5612	0.5081
13		0.4420	0	0.5580	0.5523	0.5545
14		0.4582	0	0.5418	0.5478	0.5445
15		0.4899	0	0.5101	0.5236	0.5166
16	Canteen	0.5027	0	0.4973	0.5428	0.5192
17		0.4928	0	0.5072	0.5768	0.5393
18		0.5224	0	0.4776	0.5102	0.4941
19		0.5340	0	0.4660	0.5321	0.4991
20		0.5194	0	0.4806	0.5672	0.5220
Total		10.36	0	9.64	10.8004	10.2178

<u>Constant</u>

- Distance from ground to camera : 2m
- Frame Rate: 60fps
- Distance from camera to pedestrian : 4m
- Light Intensity: 520 LUX

<u>Example of miss rate calculation</u>

Miss Rate = $FP/(TP + FP)$

Miss Rate= 0.4885/ (0.6252+0.4885) = 0.4386

Total miss rate = 10.2178

Total Rate of Error = Total Miss Rate / Total Number of test

Total Rate of Error= 10.2178/20= 0.5109

Accuracy= 1- Total Rate of Error

Accuracy= (1- 0.5109) X (100) = 0.4891%

Actual Accuracy=0.9735

Desired Accuracy=0.4891

Performance= 1- [(Actual accuracy – Desired Accuracy) / (Actual Accuracy)]

Performance= 1-[(0.9735-0.4891)/ (0.9735)]

Performance=0.5024x100%= 50.24%

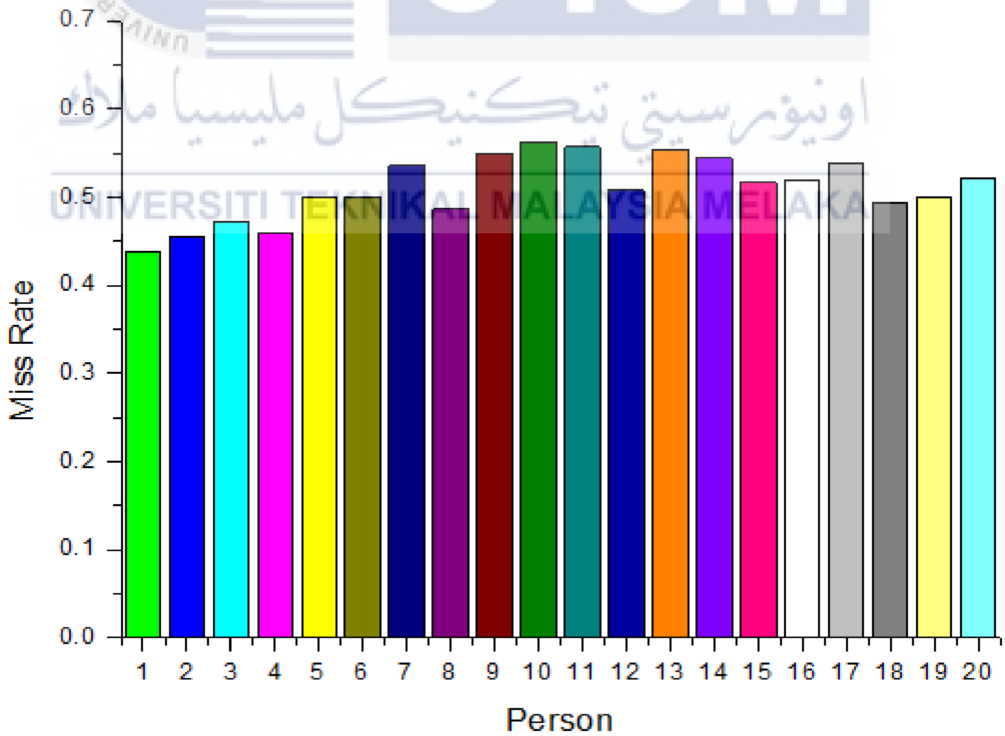


Figure 4. 9: Histogram of the miss rate when the distance is 4m

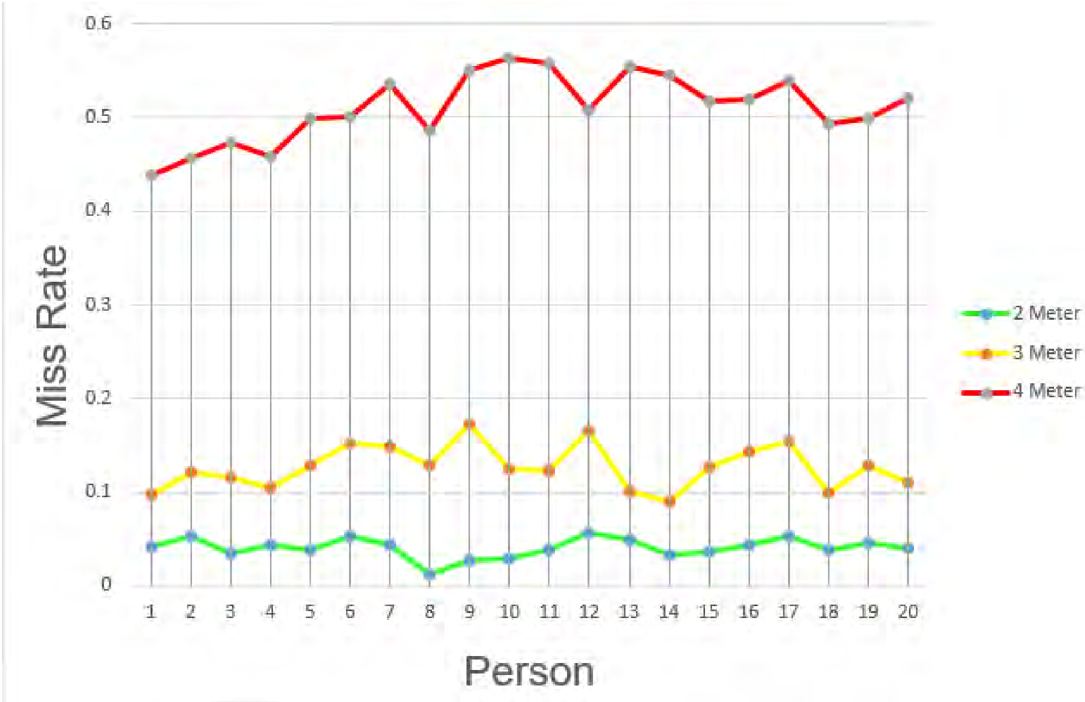


Figure 4. 10: Comparison of miss rate with different distances

4.3.1.4 Discussion from analysis of the distance between camera and the occluded human

There are three different distances between camera and pedestrian used for the experiment which is 2meter, 3meter and 4meter. This experiment is carried out to find out which distance will be very suitable for the recognition system. Sometimes the camera can also influence the recognition system because every camera has its own limit to detect specific range. The experiment has been conducted by using constant light intensity and constant frame rate of the camera (60fps). The data is collected by repeating the recognition process for 20 times for 20 different pedestrians. There are four different indoor environments which is the library, class room, house and canteen. From the experiment, influence of the distance on the system can be found out. From Table 4.2 (Result of distance with 2meter), Table 4.3 (Result of distance with 3meter) and Table 4.4 (Result of distance with 4meter), the results show that all the tests are able to recognise the pedestrians and it shows that the more the distance, the less the accuracy of recognition system. First, the accuracy of the system is 95%, when the distance is 2meter. Then, the accuracy when the distance is 3meter is 86.58%. When the system works from the distance of 4meter, the accuracy is 48.91%. The system becomes very weak in recognition when the distance increases.

Figure 4.7 shows the histogram of the miss rate when it is at a distance of 2meter from the camera to the pedestrian. The histogram shows a clear view of the error of the recognition when the distance is 2m. The highest miss rate is 5.74% and the lowest miss rate is 1.35%. Its overall mean of error is 4.08%. It is considered as low rate of error and it proves that it has very high accuracy in recognition which is 95.92%. The level of the accuracy states that it is at independent level which means it is in high accuracy and can work independently. Then the performance is 98.53%. This system perform very well when the distance is 2m and it is able to capture the pedestrian easily.

Figure 4.8 shows the histogram of the miss rate when the distance is 3meter from the camera to the pedestrian. From the histogram graph, the highest miss rate is 16.54% and the lowest miss rate is 12%. Its overall mean of error is 13.42%. It is considered as a high rate of error. So it only has the accuracy of 86.58%. The level of accuracy states that it is at frustration level which means it is in low accuracy. Then, the performance when the distance is 3meter is 88.94%. So, it still able to recognize and has a good performance.

Figure 4.9 shows the histogram of the miss rate when the distance is 4meter. The histogram shows that this system has more error when the distance is 4meter. The highest miss rate is 55.55% and the lowest miss rate is 43.86%. The overall mean of the miss rate is 51.09%. The rate of error is much higher than the accuracy of the recognition when the distance from camera to pedestrian is 4meter. The accuracy of the system is 48.91% and the performance is 52.5%. This system does not perform well in recognition when the distance is 4meter.

Figure 4.10 shows the comparison of miss rate in different distances. It clearly shows the rate of error increases as the distance increases. This may be due to the pixel of the camera or ability of the camera to get a clear view even in far range. The system can perform very accurately, if the distance of the camera is within the range of pixel. For this system, 2meter is the very best distance compared to others.

4.3.2 Analysis of Light intensity when the camera recognize the occluded human

4.3.2.1 Result of the light intensity test of dim light (320 lux)

Table 4. 5: Result of light intensity experiment

Person	Place	TP	TN	FN	FP	Miss Rate/Error
1	Library	0.8925	0	0.1075	0.2012	0.1839
2		0.9106	0	0.0894	0.2485	0.2144
3		0.8826	0	0.1174	0.1874	0.1751
4		0.8592	0	0.1408	0.1509	0.1494
5		0.8621	0	0.1379	0.1708	0.1654
6	Class Room	0.8247	0	0.1753	0.1895	0.1868
7		0.8835	0	0.1165	0.2185	0.1983
8		0.8899	0	0.1101	0.1625	0.1544
9		0.9056	0	0.0944	0.1235	0.1200
10		0.9109	0	0.0891	0.0985	0.0976
11	House	0.8327	0	0.1673	0.1825	0.1798
12		0.8358	0	0.1632	0.2456	0.2269
13		0.8577	0	0.1423	0.2238	0.2069
14		0.8945	0	0.1055	0.1586	0.1506
15		0.9086	0	0.0914	0.1260	0.1218
16	Canteen	0.8478	0	0.1522	0.1428	0.1442
17		0.8566	0	0.1434	0.2455	0.2228
18		0.8213	0	0.1787	0.1582	0.1615
19		0.8891	0	0.1109	0.1324	0.1296
20		0.8755	0	0.1245	0.1348	0.1334
Total		17.4422	0	2.5578	3.5015	3.3228

Constant

- Distance from ground to camera : 2m
- Distance from camera to pedestrian : 2m
- Frame Rate: 60fps
- Light Intensity: 320 LUX

Example of miss rate calculation

Miss Rate = $FP/(TP + FP)$

Miss Rate= 0.2012/ (0.8925+0.2012) = 0.1839

Total miss rate = 3.3228

Total Rate of Error = Total Miss Rate / Total Number of test

Total Rate of Error= 3.3228/20=0.1661

Accuracy= 1- Total Rate of Error

Accuracy= (1- 0.1661) X (100) = 83.39%

Actual Accuracy=0.9735

Desired Accuracy=0.8339

Performance= 1- [(Actual accuracy – Desired Accuracy) / (Actual Accuracy)]

Performance= 1-[(0.9735-0.8339)/ (0.9735)]

Performance=0.8566.x100%= 85.66 %

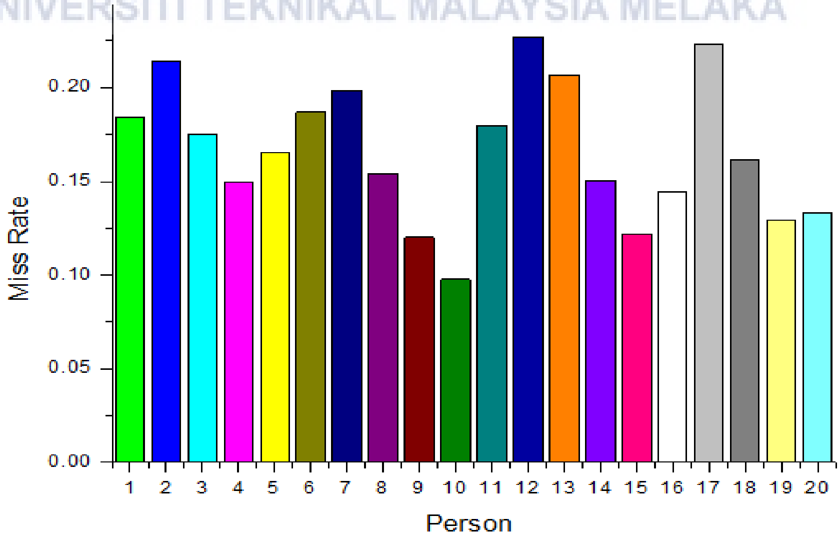


Figure 4. 11: Histogram of the miss rate when the light intensity with 320 lux

4.3.2.2 Result of the light intensity test of dim light (520 lux)

Table 4. 6: Result of light intensity experiment of (520 lux)

Person	Place	TP	TN	FN	FP	Miss Rate/Error
1	Library	0.9252	0	0.0748	0.0408	0.0422
2		0.9265	0	0.0735	0.0524	0.0535
3		0.9185	0	0.0815	0.0328	0.0345
4		0.9365	0	0.0635	0.0425	0.0434
5		0.9375	0	0.0625	0.0385	0.0394
6	Class room	0.9025	0	0.0975	0.0515	0.0540
7		0.9255	0	0.0745	0.0423	0.0437
8		0.9168	0	0.0832	0.0125	0.0135
9		0.9321	0	0.0679	0.0265	0.0276
10		0.9420	0	0.0580	0.0285	0.0294
11	House	0.9388	0	0.0612	0.0375	0.0384
12		0.9275	0	0.0725	0.0565	0.0574
13		0.9395	0	0.0605	0.0482	0.0488
14		0.9285	0	0.0715	0.0310	0.0323
15		0.9490	0	0.0510	0.0365	0.0370
16	Canteen	0.9186	0	0.0814	0.0425	0.0442
17		0.9244	0	0.0756	0.0512	0.0525
18		0.9528	0	0.0472	0.0379	0.0383
19		0.9355	0	0.0645	0.0452	0.0461
20		0.9468	0	0.0532	0.0396	0.0401
Total		18.6245	0	1.3755	0.7944	0.8163

Constant

- Distance from ground to camera : 2m
- Frame Rate: 60fps
- Distance from camera to pedestrian : 2m
- Light Intensity: 520 LUX

Example of miss rate calculation

Miss Rate = $FP/(TP + FP)$

Miss Rate= 0.0408/ (0.9252+0.0408) = 0.0422

Total miss rate = 0.8163

Total Rate of Error = Total Miss Rate / Total Number of test

Total Rate of Error= 0.8163/20= 0.0408

Accuracy= 1- Total Rate of Error

Accuracy= (1- 0.0408) X (100) = 95.92%

Actual Accuracy=0.9735

Desired Accuracy=0.9592

Performance= 1- [(Actual accuracy – Desired Accuracy) / (Actual Accuracy)]

Performance= 1-[(0.9735-0.9592)/ (0.9735)]

Performance=0.9853x100%= 98.53%

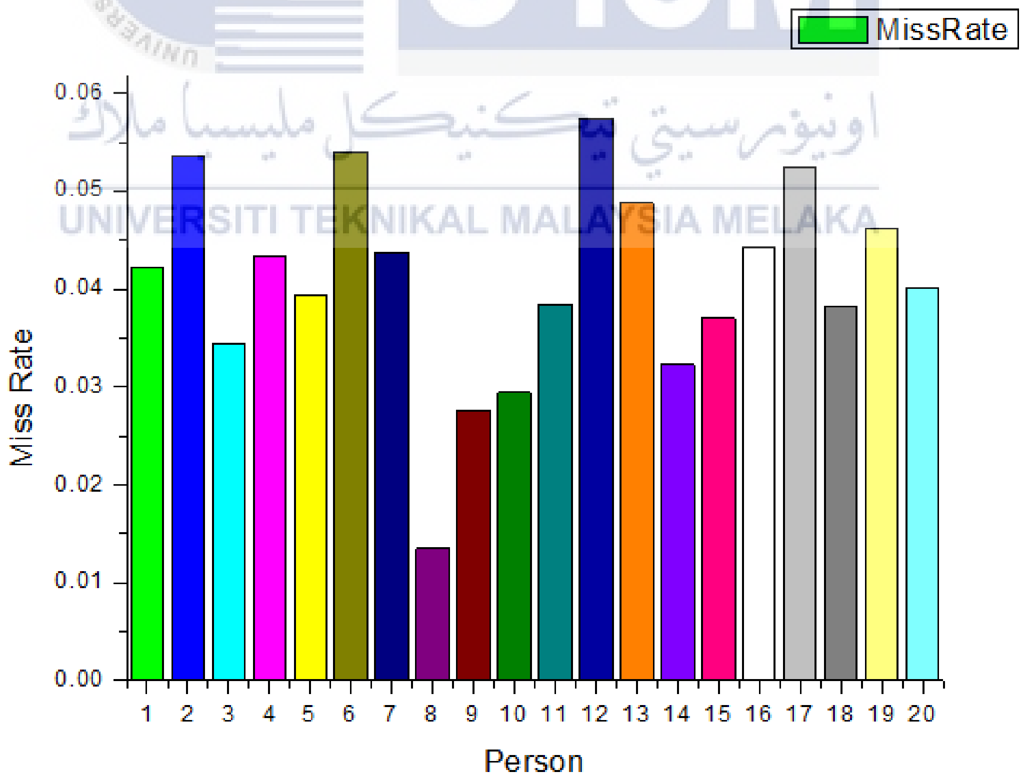


Figure 4. 12: Histogram of the miss rate when the light intensity with 520 lux

4.3.2.3 Result of the light intensity test of dim light (750 lux)

Table 4. 7: Result of light intensity experiment of (750 lux)

Person	Place	TP	TN	FN	FP	Miss Rate/Error
1		0.9326	0	0.0674	0.0789	0.0780
2		0.8824	0	0.1176	0.2359	0.2109
3	Library	0.8235	0	0.1765	0.3294	0.2857
4		0.9215	0	0.0785	0.0925	0.0912
5		0.8355	0	0.1645	0.2284	0.2147
6		0.9278	0	0.0722	0.0869	0.0856
7	Class Room	0.8160	0	0.1840	0.2162	0.2095
8		0.8354	0	0.1646	0.2245	0.2118
9		0.8769	0	0.1231	0.1678	0.1606
10		0.9406	0	0.0594	0.0825	0.0806
11		0.9120	0	0.0880	0.1276	0.1227
12		0.8177	0	0.1823	0.2354	0.2235
13	House	0.8562	0	0.1438	0.2146	0.2004
14		0.8198	0	0.1802	0.1976	0.1942
15		0.7968	0	0.2032	0.1570	0.1646
16		0.8592	0	0.1408	0.1678	0.1634
17		0.9356	0	0.0644	0.0988	0.0955
18	Canteen	0.9102	0	0.0898	0.1196	0.1161
19		0.8292	0	0.1708	0.1934	0.1891
20		0.8792	0	0.1208	0.1541	0.1491
Total		17.41	0	2.5919	3.4089	3.2472

Constant

- Distance from ground to camera : 2m
- Frame Rate: 60fps
- Distance from camera to pedestrian : 2m
- Light Intensity: 750 LUX

Example of miss rate calculation

Miss Rate = $FP/(TP + FP)$

Miss Rate= 0.0789/ (0.9326+0.0789) = 0.0780

Total miss rate = 3.2472

Total Rate of Error = Total Miss Rate / Total Number of test

Total Rate of Error= 3.2472/20= 0.1624

Accuracy= 1- Total Rate of Error

Accuracy= (1- 0.1624) X (100) = 83.76%

Actual Accuracy=0.9735

Desired Accuracy=0.8376

Performance= 1- [(Actual accuracy – Desired Accuracy) / (Actual Accuracy)]

Performance= 1-[(0.9735-0.8376)/ (0.9735)]

Performance=0.8604x100%= 86.04%

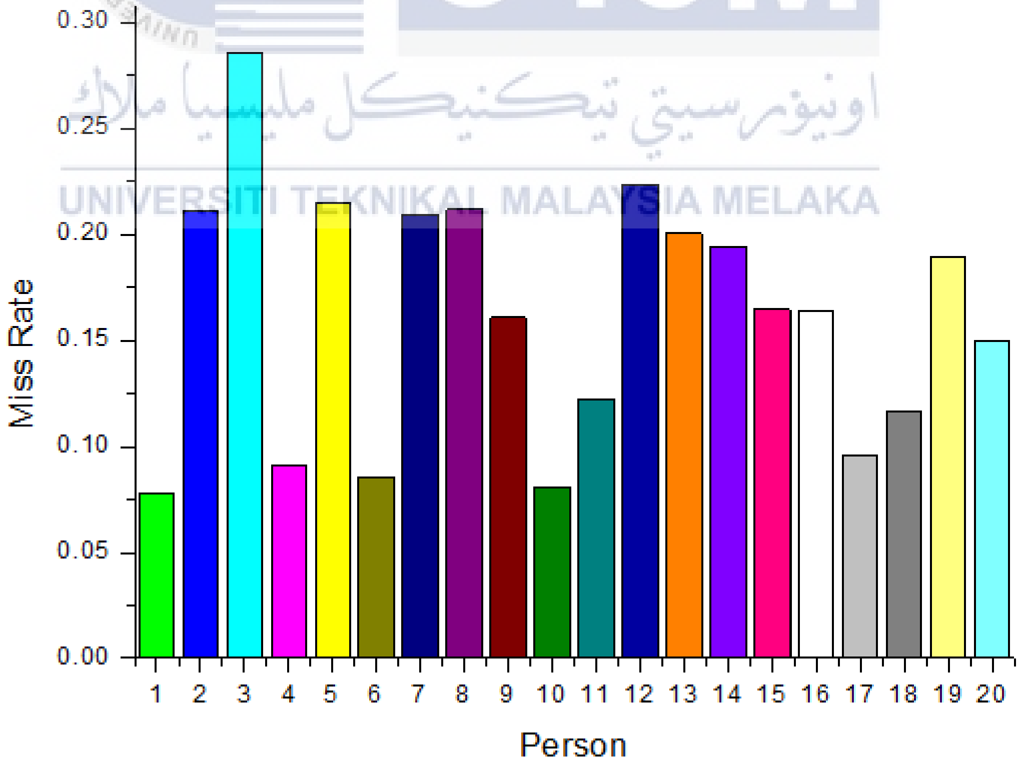


Figure 4. 13: Histogram of the miss rate when the light intensity with 750 lux

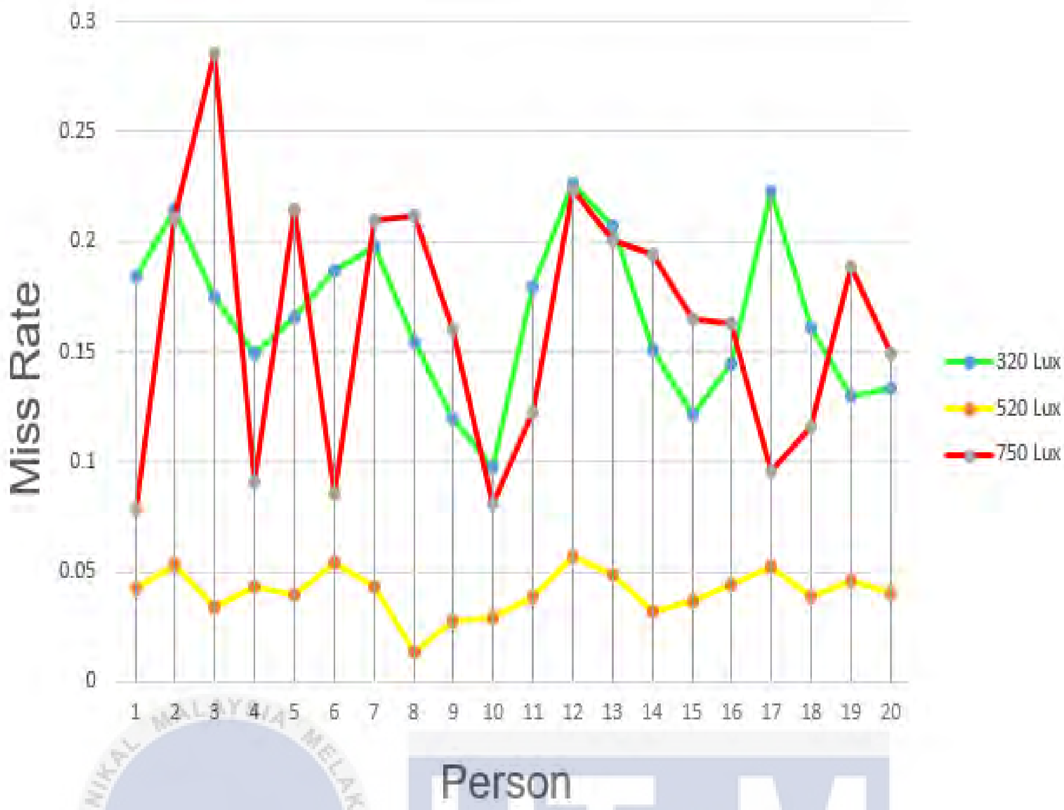


Figure 4. 14: Comparison of miss rate with different light intensity

4.3.2.4 Discussion from analysis of Light intensity when the camera recognize the occluded human

There are three different indoor light intensities used for the experiments which is dim (320 lux) light, moderate (520 lux) light and bright (750 lux) light. Indoor environments can be said as controllable environment because the light intensity can be controlled. Desired type of lighting can be decided whether it is bright, moderate or dim. The experiment is conducted by using a constant distance and constant frame rate of the camera (60fps). The data is collected by repeating the recognition process for 20 times by 20 different pedestrians. There are four different indoor environments which is the library, class room, house and canteen. The distance of the camera from the ground is 2m and the distance between the pedestrian and the camera is fixed at 2m.

From Table 4.5 (Result of Dim Light), Table 4.6 (Result of moderate light) and Table 4.7 (Result of bright light), the results show that all the tests are able to recognise the pedestrians and it also shows that light intensity can influence the accuracy of recognition. At first, the accuracy of the system is 83.39% under dim lighting (320 lux). Then, the accuracy under moderate lighting (520 lux) is 95.92%. It shows that it has high accuracy in recognition of human. When the system uses bright lighting (750 lux), the accuracy is 83.73%. The system become uncontrollable when bright light.

Figure 4.11 shows the histogram of the miss rate under the light intensity of 320 lux. The 320 lux is also called as a level of dim light. From the histogram graph, the highest miss rate is 22.69% and the lowest miss rate is 12%. The overall mean of error is 16.61%. It is considered as a high rate of error. So it only has the accuracy of 83.39%. The level of accuracy is at frustration level which means it is in low accuracy. Then, the performance is 85.66%. It is because the images captured by the camera is a bit dark and it makes the computer difficult to classify.

Figure 4.12 shows histogram of miss rate under the light intensity of 520 lux. The 520 lux is called as a level of moderate light. This type of lightings gives best result in recognition. The highest miss rate is 5.74% and the lowest miss rate is 1.35%. The overall mean of the error is 4.08%. The error is very low and the accuracy is very high (95.92%) when the lighting is moderate. The performance is 98.53%. So, it clearly shows that this system performs very well in recognition when the lighting is in moderate level.

Then, figure 4.13 shows the histogram of miss rate under the lighting of 750 lux which is a bright lighting. When the lighting is very bright, this system becomes uncontrollable which means sometimes the system is able to recognize accurately and sometimes it can’t. The highest miss rate is 28.57% and the lowest miss rate is 7.8%. Then, the overall mean of miss rate is 16.24%. It means that the system has the accuracy of 83.764%. The accuracy level is at frustration level (low). This system can perform 86.04% when the lighting is very bright. Figure 4.14 shows the comparison of miss rate with lighting levels. It clearly shows that the system is very low in error when the lighting is at moderate level. The results prove that this system is able to recognize humans if the lighting is at moderate level. The system shows low performance and in recognition when the lighting is dim and bright.

4.3.3 Analysis of accuracy and performance when the camera recognize the occluded human

In this analysis, the occlusion will classify as two category which is occlusion less than 40% and occlusion more than 40%. This 40% is just an assumption value. Less than 40 % occlusion indicates that the occlusion level is very less and also can be said as partially occluded. Then, the more than 40% occlusion means that the occlusion almost hide half or full body of human. The 40% is not a measured value. It is just an assumption value to measure the accuracy level.

4.3.3.1 Result of the Accuracy experiment (without occlusion)

Table 4. 8: Accuracy result of recognition system without occlusion

Person	Place	TP	TN	FN	FP	Miss Rate/Error
1	Library	0.9405	0	0.0595	0.0306	0.0315
2		0.9615	0	0.0385	0.0408	0.0407
3		0.9235	0	0.0765	0.0221	0.0234
4		0.9484	0	0.0516	0.0321	0.0327
5		0.9555	0	0.0445	0.0301	0.0305
6	Class room	0.9485	0	0.0515	0.0461	0.0463
7		0.9755	0	0.0245	0.0282	0.0281
8		0.9245	0	0.0755	0.0345	0.0359
9		0.9384	0	0.0616	0.0145	0.0152
10		0.9186	0	0.0814	0.0125	0.0134
11	House	0.9856	0	0.0144	0.0197	0.0196
12		0.9756	0	0.0244	0.0250	0.0249
13		0.9748	0	0.0252	0.0280	0.0279
14		0.9439	0	0.0561	0.0310	0.0318
15		0.9542	0	0.0458	0.0355	0.0358
16	Canteen	0.9651	0	0.0349	0.0125	0.0128
17		0.9732	0	0.0268	0.0155	0.0157
18		0.9845	0	0.0155	0.0206	0.0205
19		0.9902	0	0.0098	0.0185	0.0183
20		0.9839	0	0.0161	0.0245	0.0243
Total		19.1659	0	0.8341	0.5223	0.5293

Constant

- Distance from ground to camera : 2m
- Frame Rate: 60fps
- Light Intensity: 520 LUX

Example of miss rate calculation

Miss Rate = $FP/(TP + FP)$

Miss Rate= 0.0306/ (0.9405+0.0306) = 0.0315

Total miss rate = 0.5293

Total Rate of Error = Total Miss Rate / Total Number of test

Total Rate of Error= 0.5293/20= 0.0265

Accuracy= 1- Total Rate of Error

Accuracy= (1- 0.0265) X (100) = 97.35%

Actual Accuracy=0.9735

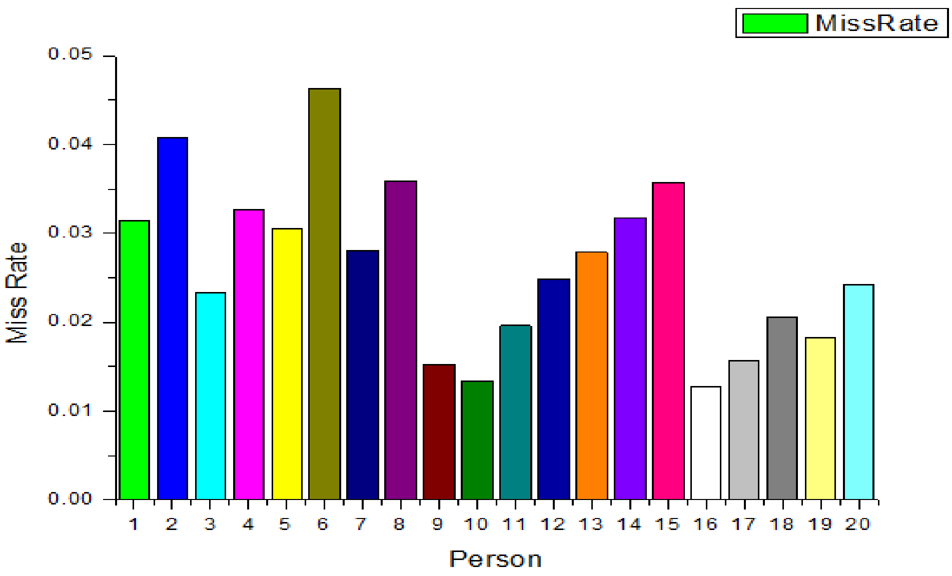


Figure 4. 15: Histogram of the miss rate without occlusion

4.3.3.2 Result of the Accuracy experiment (With Occlusion of less than 40%)

Table 4. 9: Accuracy result of recognition system with Occlusion of less than 40%

Person	Place	TP	TN	FN	FP	Miss Rate/Error
1	Library	0.9252	0	0.0748	0.0408	0.0422
2		0.9265	0	0.0735	0.0524	0.0535
3		0.9185	0	0.0815	0.0328	0.0345
4		0.9365	0	0.0635	0.0425	0.0434
5		0.9375	0	0.0625	0.0385	0.0394
6	Class room	0.9025	0	0.0975	0.0515	0.0540
7		0.9255	0	0.0745	0.0423	0.0437
8		0.9168	0	0.0832	0.0125	0.0135
9		0.9321	0	0.0679	0.0265	0.0276
10		0.9420	0	0.0580	0.0285	0.0294
11	House	0.9388	0	0.0612	0.0375	0.0384
12		0.9275	0	0.0725	0.0565	0.0574
13		0.9395	0	0.0605	0.0482	0.0488
14		0.9285	0	0.0715	0.0310	0.0323
15		0.9490	0	0.0510	0.0365	0.0370
16	Canteen	0.9186	0	0.0814	0.0425	0.0442
17		0.9244	0	0.0756	0.0512	0.0525
18		0.9528	0	0.0472	0.0379	0.0383
19		0.9355	0	0.0645	0.0452	0.0461
20		0.9468	0	0.0532	0.0396	0.0401
Total		18.6245	0	1.3755	0.7944	0.8163

Constant

- Distance from ground to camera : 2m
- Frame Rate: 60fps
- Light Intensity: 520 LUX

Example of miss rate calculation

Miss Rate = $FP/(TP + FP)$

Miss Rate= 0.0408/ (0.9252+0.0408) = 0.0422

Total miss rate = 0.8163

Total Rate of Error = Total Miss Rate / Total Number of test

Total Rate of Error= 0.8163/20= 0.0408

Accuracy= 1- Total Rate of Error

Accuracy= (1- 0.0408) X (100) = 95.92%

Actual Accuracy=0.9735

Desired Accuracy=0.9592

Performance= 1- [(Actual accuracy – Desired Accuracy) / (Actual Accuracy)]

Performance= 1-[(0.9735-0.9592)/ (0.9735)]

Performance=0.9853x100%= 98.53%

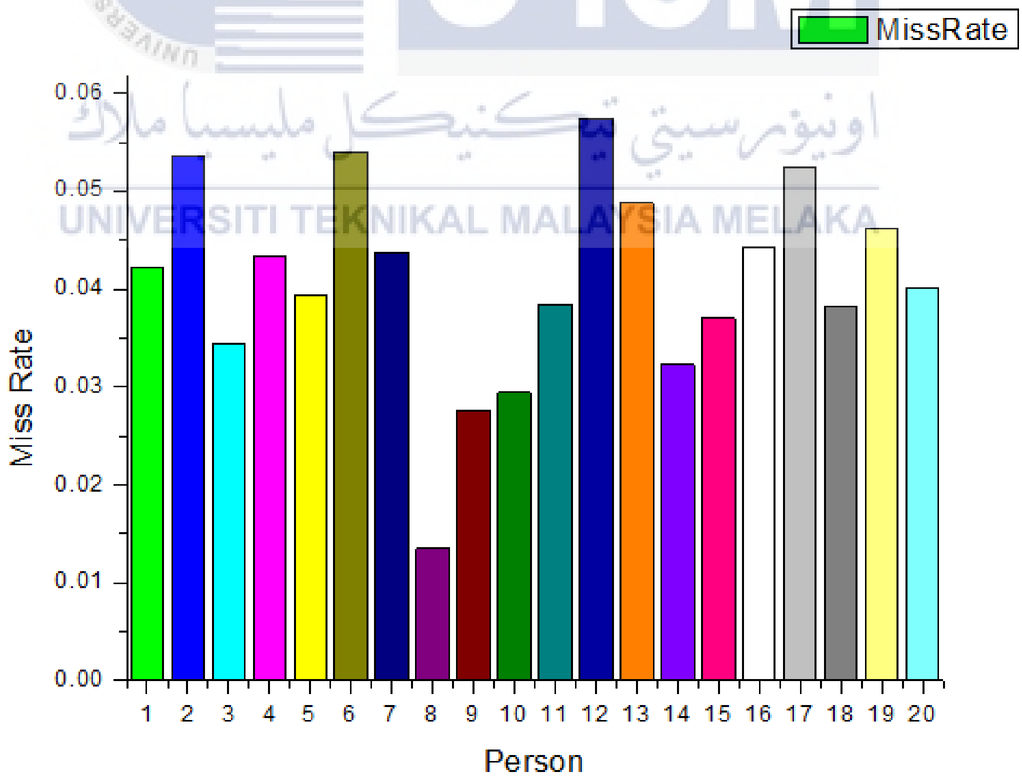


Figure 4. 16: Histogram of the miss rate with occlusion less than 40%

4.3.3.3 Result of the Accuracy experiment (With Occlusion of more than 40%)

Table 4. 10: Accuracy result of recognition system with occlusion of more than 40%

Person	Place	TP	TN	FN	FP	Miss Rate/Error
1	Library	0.6425	0	0.3575	0.3685	0.3645
2		0.5725	0	0.4275	0.4175	0.4217
3		0.6236	0	0.3764	0.3856	0.3821
4		0.5475	0	0.4525	0.4765	0.4653
5		0.5535	0	0.4465	0.4685	0.4584
6	Class room	0.5215	0	0.4785	0.5183	0.4985
7		0.4856	0	0.5144	0.5255	0.4803
8		0.4987	0	0.5022	0.5355	0.5182
9		0.5196	0	0.4804	0.5155	0.4980
10		0.5595	0	0.4405	0.4855	0.4646
11	House	0.4294	0	0.5706	0.5854	0.5769
12		0.5872	0	0.4128	0.4235	0.4190
13		0.6120	0	0.3880	0.4120	0.4023
14		0.5025	0	0.4975	0.5125	0.5051
15		0.4321	0	0.5679	0.5755	0.5712
16	Canteen	0.5118	0	0.4882	0.3855	0.4296
17		0.4965	0	0.5035	0.4558	0.4786
18		0.3695	0	0.6305	0.6825	0.6488
19		0.5429	0	0.4571	0.4675	0.4627
20		0.5235	0	0.4765	0.3885	0.4250
Total		10.531	0	9.469	9.5856	9.4708

Constant

- Distance from ground to camera : 2m
- Frame Rate: 60fps
- Light Intensity: 520 LUX

Example of miss rate calculation

Miss Rate = $FP/(TP + FP)$

Miss Rate= 0.3685/ (0.6425+0.3685) = 0.3645

Total miss rate = 9.4708

Total Rate of Error = Total Miss Rate / Total Number of test

Total Rate of Error= 9.4708/20=0.4735

Accuracy= 1- Total Rate of Error

Accuracy= (1- 0.4735) X (100) = 52.95%

Actual Accuracy=0.9735

Desired Accuracy=0.5295

Performance= 1- [(Actual accuracy – Desired Accuracy) / (Actual Accuracy)]

Performance= 1-[(0.9735-0.5295)/ (0.9735)]

Performance=0.4560.x100%= 45.6%

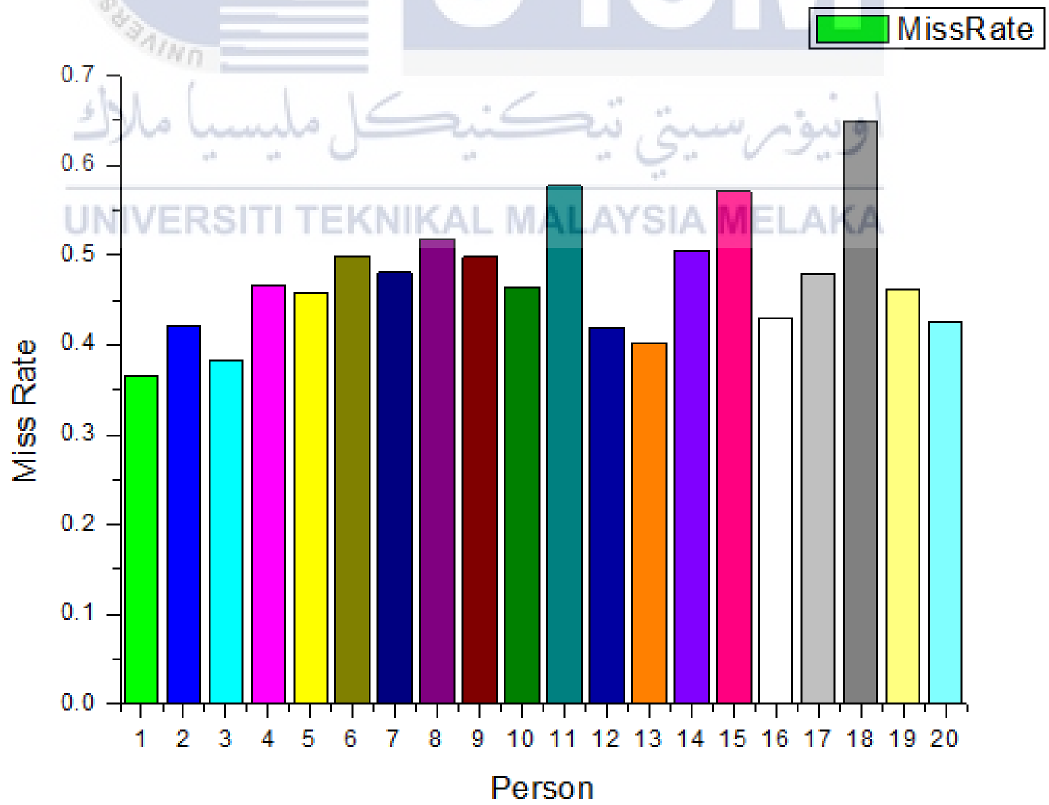


Figure 4. 17: Histogram of the miss rate with occlusion more than 40%

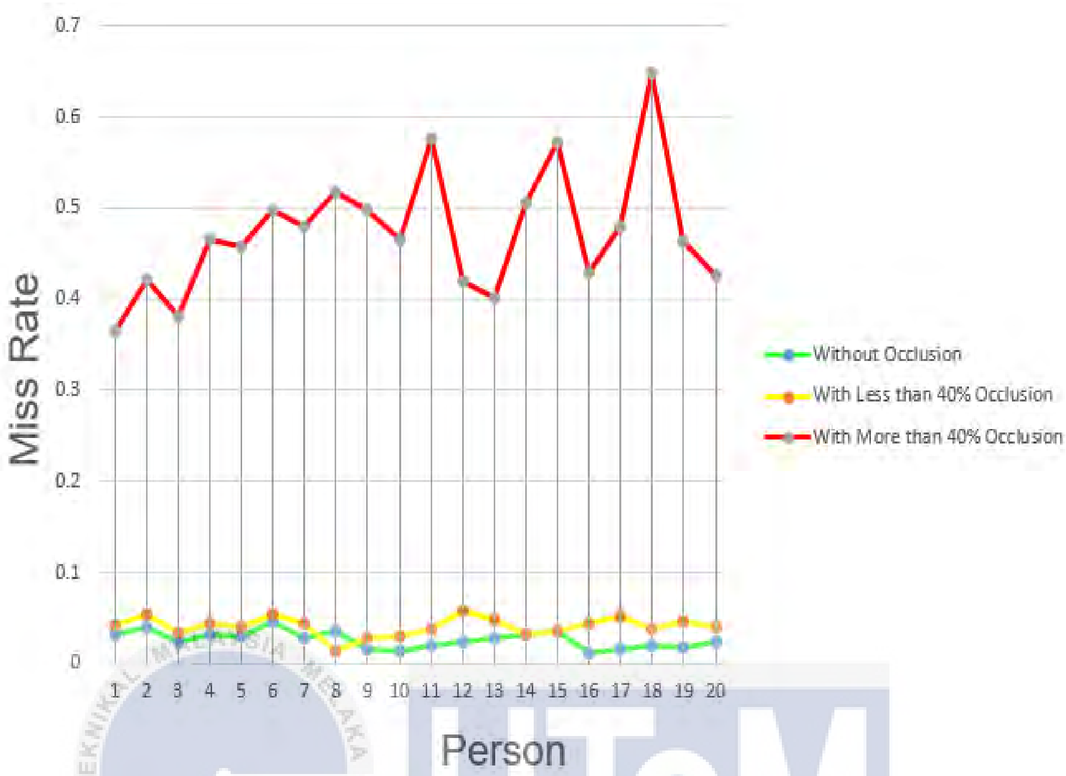


Figure 4. 18: Histogram of comparison of overall miss rate

Figure 4. 19: Pedestrian without any occlusion

Figure 4. 20: The pedestrian with less than 40% of occlusion

Figure 4. 21: The pedestrian with more than 40% occlusion

4.3.3.4 Discussion of accuracy analysis

The accuracy experiment has been conducted to check the reliability of the human pedestrian recognition system. This experiment is conducted to find out the accuracy of the system when recognizing an occluded person. It has been conducted with constant distance, constant light intensity (520lux) and constant frame rate of the camera (60fps). The data is collected by repeating the recognition process for 20 times for 20 different pedestrians. This repeating process enables us to find the error of the system. The experiment is conducted in indoor environment. There are four different indoor environments which is the library, class room, house and canteen. The distance of the camera from the ground is 2m and the distance between pedestrians and the camera is fixed at 2m.

From Table 4.8 (Accuracy test without occlusion), Table 4.9 (Accuracy test with occlusion of less than 40%) and Table 4.10 (Accuracy test with occlusion more than 40%), the results show that all the tests are able to recognise the pedestrians and it also shows that the more the occlusion is, the lesser its accuracy. At first, the accuracy of the system is 97.35% when there is no occlusion. It is also the actual accuracy of the system. The result proved that the system is very accurate in human recognition. Then, the accuracy when there is occlusion (with less than 40%) is 95.92%. This shows that it has high accuracy in recognition of the humans when partially occluded. When there is occlusion more than 40%, the accuracy is 52.95%. It means that the system has very low accuracy when more than 40% of occlusion happens.

Figure 4.15 shows the histogram of the miss rate without occlusion. The miss rate means error of the system. The histogram shows that the highest miss rate is 0.0463 which means 4.63% and the lowest miss rate is 0.0128 which is 1.28%. Then, the mean of the miss rate is 0.0265 which is 2.65%. So the result shows that this system has a very low rate of error in recognition when there is no occlusion. It has only overall of 2.65% error in recognition. It means that this system is able to recognize human very accurately as 97.35%. This 97.35% is also the actual value of accuracy. The level of the accuracy of this system is at independent level at which it has the ability to recognize human independently without any instruction required.

Then, figure 4.16 shows the histogram of the miss rate with occlusion of less than 40%. It means that the pedestrian that passes by the camera is partially occluded. The results are quite similar to the accuracy of without occlusion which is the highest miss rate is 5.74% and lowest miss rate is 1.35%. Then, the overall mean of the miss rate is 4.08%. It means when occlusion happens, the rate of error is about 4.08% and the accuracy is 95.42%. So, the level of the accuracy is at independent level. It also has the ability to recognize the human independently when the occlusion happens less than 40%. The results show that it has a high accuracy and also a high performance. The performance when occlusion of less than 40% happens is 98.53%. It proves that this system performs very well in human recognition even when it is partially occluded.

Figure 4.17 shows the histogram of miss rate with occlusion of more than 40%. The histogram shows that this system has more error when the occlusion is more than 40%. The highest miss rate is 57.69% and the lowest miss rate is 36.45%. The overall mean of the miss rate is 47.35%. The rate of error is very high when the occlusion is more than 40%. Occlusion more than 40% means half of the human body is hidden by objects or anything. When this kind of occlusion happens, the accuracy of the system is 52.65 % and the level of the accuracy is at frustration level. The frustration level means that this system is considered as weak in detection when the occlusion is more than 40%. Then, the performance of this system is 45.6% which means that this system does not perform well in recognition when the occlusion is more than 40%.

Figure 4.18 shows histogram of comparison of miss rate. It clearly shows that when the occlusion is more than 40%, the rate of error becomes higher. It means the recognition will be less accurate. So the rate of occlusion can be a problem for a human recognition system. The results and data prove that the more the occlusion rate, the less the accuracy. This problem can be solved if more data is used to train the system. Sometimes the light intensity also can influence the accuracy of the recognition. This pedestrian recognition system can work accurately if the pedestrian is partially occluded which is in 95.42% of accuracy and 98.53% of performance. If the pedestrians are occluded by more than 40%, the system will be less accurate which is in 52.65% of accuracy and 45.6% of performance.

4.3.4 Result Conclusion

In this chapter, all the results are shown and discussed in detail. All the results are obtained from the experiments and tests done. So from that, the data is tabulated into table and plotted into graph and analysed. All the experiments are focused on the occluded pedestrian that passes by it. The results show that this human recognition system performs very well and classify the pedestrians accurately.

Accuracy and performance experiments are very important for any kind of systems. It will clearly show on how well the system can perform in real life situations. This system is very accurate in recognition for about 97.35%. When occlusion happens, this system becomes less accurate in recognition. The more the occlusion, the lesser the accuracy. When recognizing the pedestrian that is occluded less than 40%, this system is still accurate in recognition but when it recognizes the pedestrian whom are occluded more than 40%, this system becomes very weak in recognizing the pedestrian. So, this problem can be handled by importing more data to train the system. The more the data, the higher the accuracy.

Then, the system is tested with light intensity. Moderate light gives the best result for recognition. Light intensity is the most important factor that must be handled for any kind of image processing technique. The accuracy and performance are very high for moderate light compared to bright and dim light. Then, the distance experiment is done to find out the suitable distance for this system. The results show that the system's accuracy and performance decreases as the distance increases. This may be due to the ability of the camera. This problem can be solved by using a camera that has a high range of pixel.

Hence, from the results it can be seen that the occlusion, distance and light intensity can influence the accuracy and performance of an image processing system. It requires more data and proper research to handle all those problems. Camera is the eye of any image processing system. It must be in high range of pixel so that it can capture the image very clearly.

CHAPTER 5

5.1 Conclusion

The first objective of this research is to recognize human pedestrian by using Haar Cascade and Convolutional Neural Network methods. This objective is achieved and the system is also able to recognize accurately. The preferred methods and techniques are selected after deep study is conducted on the journal and other sources. The second objective is to analyse the performance of the human recognition system when occlusion happens. The second objective has also been achieved after several tests and experiments. The whole analysis only focuses on the occluded human. The data obtained from the experiments of the system is analysed in order to achieve the objective. From the data, histogram graph has been plotted and several conclusion and discussion is made.

Image processing in real time video is a difficult task in computer vision system. Computers see the images in different way which is in number of matrix. It requires more data to understand and classify the contents of an image. To train the computer to recognize an object, the computer itself must be in a high specification which is with high ram and fast processor. It allows the videos to be smoother and can train the datasets easily in a fast way.

Haar cascade can be said as the best technique for fast and accurate detection of the human pedestrian. Haar cascade can be trained very easily to detect humans by using positive and negative images. The Convolutional Neural Network is a best classification method which is used in deep learning technique. It is able to classify the class of the human pedestrian automatically and accurately. The Keras API is chosen for this system. It is one of the easiest API to create human models compared to the others. Keras API is also very suitable for beginners to learn about deep learning and to solve real world problems.

This system is more suitable for indoor environment because the indoor environment is controllable. Then, this system is also very accurate even when the pedestrian is partially occluded but if the occlusion increases the accuracy might decrease. Besides, moderate lighting is the most suitable aspect for this system. The results show that moderate lighting gives highest accuracy in recognition. For the computer vision system, light intensity will be a big challenge for it to handle. Then, this system is also suitable in recognizing human in the range of 2 to 3meter from the camera. The distance problem may be due to the pixels of the camera or ability of the camera to get a clear view even when in it is in a far range. All the conclusions are made based on the experimental results. From all of the experiments and tests that are done, it can be said that all the objectives of this project are achieved successfully.

5.2 Recommendation

While doing research and development of human recognition system, there are some errors and difficulties to complete it. Video processing is a difficult task in computer vision system because it needs to process the scene frame by frame. To have a better and faster system, the system must be updated to faster frame rate per second (fps). Therefore the system can work very well in a fast way. Then, high specification computer can improve the performance of the system.

To get better results, the camera itself must have a great ability such as high pixels and low distortion in lens. To make the image clear from any noise, more filters should be added on the image. This system has the ability to recognize humans in real time video stream. So, maybe this system can be improved by adding some features such as online human recognition. Besides, maybe it can also email the user if human intruder is recognized when passing by the camera. Then, this recognition system can also be updated by automatically recognizing and capturing humans and storing the data to a folder according to the time, date and place.

CHAPTER 6

REFERENCES

[1] "Panduan Belajar Cara Mempertahankan Diri Untuk Keselamatan", Statistik Indeks Jenayah. Available : http://myrujukan.com/belajar

[2] Utusan Malaysia. (2016, January 4)." 659 kanak-kanak masih hilang" [Online]. Available: http://www.utusan.com.my/berita/jenayah/659-kanak-kanak-masih-hilang-1.176198

[3] "Introduction to Machine Vision" , COGNEX. Available: http://www.cognex.com/global/DownloadAsset.aspx?id=13853

[4] Danny Shapiro,(February 24,2015)," How Deep Learning Will Accelerate self- Driving cars".Available: https://blogs.nvidia.com/blog/2015/02/24/deep-learning-drive/

[5] "Cascade Classifier for People Detection in Surveillance Videos", VIVA lab. Available: http://www.site.uottawa.ca/~laganier/projects/peopledetect

[6] R. Labudzki, S. Legutko, Poznan University of Technology, Applications of Machine Vision, Faculty of Mechanical Engineering and management,2013

[7] Luca Scagliarini, Marco Varone ,Expert System, “What is machine learning,”.Available : http://www.expertsystem.com/machine-learning-definition/

[8] Muhamed Hussein, Wael Abt-Almaged, Yang Ran, Larry Davis, “Real time human detection, tracking and verification in uncontrolled camera motion environment,” IOSR Journal Of Computer Engineering,2013

[9] M. Citra, M. Kalaiselvi Geetha, L.Menaka, “Occlusion and abandoned object detection for surveillance applications,” International Journal of Computer Applications Technology and Research, pp. 708-713,2013

[10] Vandit Gajjar, Ayesha Gurnani, Yash Kandhediya,” Human Tracking in video Surveillance”, L.D College of Engineering Ahmedabad,2017

[11] Prithviraj Banarjee, Somnath sengupta,” Human motion detection and tracking for video surveillance”, Indian Institute of Technology, pp.88-92,2008

[12] Lim Eng Aik, Zarita Zainuddin,” Real-Time people counting system using curve analysis method”, International Journal of Computer and Electrical Engineering, pp. 77-83, 2009

[13] Suresh D S , Lavanya M P,” Motion detection and tracking using background subtraction and consecutive frame difference method”, Channabasaveshwara Institute of Technology, pp.16-22, 2014

[14] Vocal Complete solution," Histogram of Oriented Gradients (HOG) for Object Detection"[online] Available: https://www.vocal.com/video/histogram-of-oriented-gradients-hog-for-object-detection/,2012

[15] Dataaspirant, "SVM classifier, Introduction to support vector machine algorithm",(january 13, 2017 [online] available: http://dataaspirant.com/2017/01/13/support-vector-machine-algorithm/

[16] Dr. Andrew G. Barto, Dept. of Computer Science, University of Massachuestts - Amherst" Temporal difference learning", pp.15-18,2007

[17] Anelia Angelova, Alex krizhevsky, Vincent vanhoute, Abhijit ogale, Dave fergusan," Real time pedestrian detection with deep network cascades", 1600 Amphitheatre Parkway, pp. 12-32,2015.

[18] Saloni Navgire, Arshiyana khan, Jyoti Thorat, S.N Kulkarni, "Human detection through processing video captured by camera module interfaced using raspberry and OpenCV", N.D, 2012

[19] Alok.k. Singh Kushwala, Chandra Mani Sharma, " Automatic multiple human detection and tracking for visual surveillance system", GLA University, Mathura, INDIA,2012

[20] Er. Novjot Kaur, Er. Yadwinder kaur, "Object classifications technique using machine learning models", Chandigarh Group of Colleges, Gharuan, Mohali, pp. 170-174, 2014

[21] Shreyank N Gowda," Human activity recognition using combinational deep belief networks", Indian Institute Of Science Bangalore, India, pp. 120-125, 2014

[22] Guang Shu, " Human Detection, Tracking and Segmentation In Surveillance Video", M.S. Shanghai Jiaotong University,2009.

[23] Y.Benezeth, P.Jodoin, B.Emile, H.Laurent, C.Rosenberger Society of Photo Optical Instrumentation Engineer,2010

[24] "Deep learning methods and frameworks", VIVA research lab. [online] Available: https://project.inria.fr/deeplearning/files/2016/05/DLFrameworks

[25] S.Ikemura, H.Fujiyoshi, " Comparison of the software" Dept. of computer Science, Chubu University, 2013.

[26] "Deep Learning Basics Neural Networks Backpropagation",alexminnar research lab. [Online] Available: http://alexminnaar.com/deep-learning-basics-neural-networks

APPENDICES

```
rom keras.preprocessing.image import img_to_array
from keras.models import load_model
from imutils.video import VideoStream
from detector import BodyDetector
import argparse
import numpy as np
import imutils
import cv2



ap=argparse.ArgumentParser()
ap.add_argument("-b","--body",required=True,help="path to the body
detector")
args=vars(ap.parse_args())

bd=BodyDetector(args["body"])

MODEL_PATH = "sasi2.model"

print("[INFO] starting video stream...")
vs = VideoStream(src=0).start()


# loop over the frames from the video stream
while True:

    # to have a maximum width of 400 pixels
    frame = vs.read()
    frame = imutils.resize(frame, width=400)
    (h, w) = frame.shape[:2]


    Rects = bd.detect(frame)
    frameclone = frame.copy()

    for (fX,fY,fW,fH) in Rects:


        # extract the data
        (giri,sasi,unknown) = model.predict(image)[0]


        if (giri >sasi) & (giri>unknown):
            label = "Giri"
            proba =  giri

        if (sasi >giri) & (sasi>unknown):
            label="Sasi"
            proba=sasi


        if (unknown>giri) & (unknown>sasi):
            label = "Unknown"
            proba = unknown


        label1 = "{}: {:.2f}%".format(label, proba * 100)
        frame = cv2.putText(frameclone, label1, (fX,fY-30),
                            cv2.FONT_HERSHEY_SIMPLEX, 0.7, (0, 255, 0), 2)
        cv2.rectangle(frameclone,(fX,fY),(fX+fW,fY+600),(0,0,255),2)
```

```
    cv2.imshow("Frame", frameclone)
    key = cv2.waitKey(1) & 0xFF

    if key == ord("q"):
        break

print("[INFO] cleaning up...")
cv2.destroyAllWindows()
vs.stop()
```

Training

```
from keras.models import Sequential
from keras.layers.normalization import BatchNormalization
from keras.layers.convolutional import Conv2D
from keras.layers.convolutional import MaxPooling2D
from keras import backend as K


class SmallerVGGNet:
   @staticmethod
   def build(width, height, depth, classes):
      # initialize the model along with the input shape to be
      # "channels last" and the channels dimension itself
      model = Sequential()
      inputShape = (height, width, depth)
      chanDim = -1

      # if we are using "channels first", update the input shape
      # and channels dimension
      if K.image_data_format() == "channels_first":
         inputShape = (depth, height, width)
         chanDim = 1

      # CONV => RELU => POOL
      model.add(Conv2D(32, (3, 3), padding="same",
         input_shape=inputShape))
      model.add(Activation("relu"))
      model.add(BatchNormalization(axis=chanDim))
      model.add(MaxPooling2D(pool_size=(3, 3)))
      model.add(Dropout(0.25))

      # (CONV => RELU) * 2 => POOL
      model.add(Conv2D(64, (3, 3), padding="same",
         input_shape=inputShape))
      model.add(Activation("relu"))
      model.add(BatchNormalization(axis=chanDim))
      model.add(Conv2D(64, (3, 3), padding="same",
      model.add(Dropout(0.5))

      # softmax classifier
      model.add(Dense(classes))
      model.add(Activation("softmax"))

      # return the constructed network architecture
      return model
```

Project Activity	Sept-17			Oct-17				Nov-17				Dec-17			
FYP 1	Weeks														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Title Discussion(briefing) and Selection															
Journal finding and Literature Review															
Set the objectives and scope															
Selection of methods															
Experiments using detection and recognition method															
Preliminary Result and Discussion															
First Draft of Report Compiling															
Submission of First Draft Report															
FYP 1 Presentation															
Submission of Final FYP 1 Report															

The Gantt chart for FYP 1

Project Activity	Feb-18			Mar-18				Apr-18				May-18			
FYP 2	Weeks														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Meeting with supervisor															
Hardware and software development • Train negative and positive images • Prepare dataset • Train convolutional neural network • Run program															
Experiment • Test the accuracy of human detection and classification when occlusion															
Result and Discussion															
First Draft of Report Compiling															
Submission of First Draft Report															
FYP 2 Presentation															
Submission of Final FYP 2 Report															

The Gantt chart for FYP 2

sasinthiran fyp

ORIGINALITY REPORT

6%	3%	3%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	1%
2	Submitted to Universiti Teknikal Malaysia Melaka Student Paper	1%
3	Submitted to Universiti Malaysia Perlis Student Paper	<1%
4	www.cs.umd.edu Internet Source	<1%
5	espace.curtin.edu.au Internet Source	<1%

<1%