AN OBJECT FINDER FOR THE VISUALLY IMPAIRED

BY

CHOO YONG QUAN

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF INFORMATION SYSTEMS (HONOURS) INFORMATION SYSTEMS ENGINEERING

Faculty of Information and Communication Technology

(Kampar Campus)

JAN 2022

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: AN OBJECT FINDER FOR THE VISUALLY IMPAIRED

Academic Session: JANUARY 2022

I CHOO YONG QUAN

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

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

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

Verified by,

_________________________ _________________________

(Author's signature) (Supervisor's signature)

Address:

39, Jalan Putri Jaya 19,

Taman Putri Jaya,

43200 Cheras Selangor

Leung Kar Hang

Supervisor's name

Date: 22 April 2022

Date: 22 April 2022

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 22 April 2022

SUBMISSION OF FINAL YEAR PROJECT /DISSERTATION/THESIS

It is hereby certified that ______***Choo Yong Quan***____________________ (ID No: __***1803979***__) has completed this final year project entitled "_*An Object Finder for the Visually Impaired*_" under the supervision of _____***Leung Kar Hang***___________ (Supervisor) from the Department of Computer Science , Faculty of Information and Communication Technology , and _***Cik Norazira Binti A Jalil***_________ (Co-Supervisor)* from the Department of Information Systems, Faculty of Information and Communication Technology.

I understand that University will upload softcopy of my final year project / dissertation/ thesis* in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

____

(*Choo Yong Quan*)

DECLARATION OF ORIGINALITY

I declare that this report entitled "**AN OBJECT FINDER FOR THE VISUALLY IMPAIRED**" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : ____________________

Name : Choo Yong Quan

Date : 22 April 2022

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude and appreciations to my supervisor, Prof. Dr. Leung Kar Hang for his supervision and suggestion in guiding me toward the end of the project. The project would be impossible to complete without his precious suggestions and recommendations for my consideration in the conducting the project. His professional experiences constantly giving me new insights in the project development and achievement. Prof Leung also used his spare time to make meetings regularly with me in order to keep track and guide me throughout the project. Thank you once again to Prof Leung to your endless patience and encouragement all the time.

Secondly, I sincerely thanks to my friends for supporting me through word of encouragement to motivate me during the final semester. They have offered me snacks and drink to relieve my stress while completing my project. Thank you to them for bringing me joy and reduce my tension while accomplishing the project.

Lastly, I heartily thanks to my parents and family in providing me love and support through the difficult time. Their encouragement helps me to bring out of the depressed time and boost my confidence throughout the process.

ABSTRACT

Vision loss or vision impairment are impactful to human. The visually impaired people lost the sense of vision for them to determine the surrounding environment precisely. They are having a hard time in their life and require assistance for to carry on their daily life. In this project, an assistive solution is proposed to help the people with visual impairment to identify and even locate objects accurately to pick up the object. This project aims to develop an object finder application that can help the visually impaired people to easily locate and identify common objects or daily essentials in the indoor environment. The system will implement an object detection module developed using the pre-trained YOLO model to detects object in a single-shot convolutional neural network for real time detection from the images or video. The YOLO model is high speed and high accuracy to archive high responsiveness to the users. The system will output the detection result by label the bounding boxes of the objects in the visual data to enable the system to compute the relative direction of the object from the user's hand in the next stage. To implement hand detection, the background frame of the video is extracted and will be used to perform background subtraction with the subsequent frame. The system will then compute the direction of the object relative to the user's hand. The overlap of the user's hand with the objects in the visual data will allow the system to trigger a notification to the user to pick up the object.

TABLE OF CONTENTS

TITLE PAGE **i**

REPORT STATUS DECLARATION FORM **ii**

FYP THESIS SUBMISSION FORM **iii**

DECLARATION OF ORIGINALITY **iv**

ACKNOWLEDGEMENTS **v**

ABSTRACT **vi**

TABLE OF CONTENTS **vii**

LIST OF FIGURES **x**

LIST OF TABLES **xii**

LIST OF SYMBOLS **xiii**

LIST OF ABBREVIATIONS **xiv**

CHAPTER 1 INTRODUCTION **1**

1.1 Problem Statement and Motivation 1

1.2 Project Scope 2

1.3 Project Objectives 2

1.4 Impact, Significance and Contribution 3

1.5 Background Information 4

1.6 Report Organization 7

CHAPTER 2 LITERATURE REVIEW **8**

2.1 Review on Previous Works 8

2.1.1 Kinect Cane System [8] 8

2.1.2 Smart Cap System [9] 11

2.1.3 Web-based Object Detection and Sound Feedback System [10] 13

2.1.4 Assistive Suitcase System, Bbeep [11] 15

2.1.5 Wearable Vision-based System [12] 17

2.2 Table of Comparison Between the Existing Works 19

CHAPTER 3 SYSTEM METHODOLOGY/APPROACH 20

3.1 Design Specifications 20

3.1.1 Methodologies and General Work Procedures 20

3.1.2 Tools to Use 21

3.1.3 User Requirements 22

3.2 System Design Diagram 22

3.2.1 Use Case Diagram 22

3.2.2 Activity Diagram 23

3.3 Verification Plan 24

CHAPTER 4 SYSTEM IMPLEMENTATION 25

4.1 Overview 25

4.2 Background Subtraction 25

4.3 Object Detection 26

4.3.1 Setup Neural Network for YOLO Model 26

4.3.2 OpenCV blobFromImage Preprocessing 27

4.3.3 Evaluate Detection Results 27

4.4 Hand detection 28

4.4.1 Skin Colour Detection using YCbCr Colour Space 29

4.4.2 Hand Contour and Finger Detection 30

4.5 Compute Direction 32

CHAPTER 5 SYSTEM EVALUATION AND DISCUSSION 35

5.1 System Testing and Performance Metrics 35

5.2 Testing Setup and Result 35

5.2.1 Testing on Different Brightness by Manipulating the Light Condition 35
5.2.2 Testing on Different Types of Background 39
5.2.3 Testing on Varying Number of Objects Class 45
5.2.4 Testing on Changing Targeted Object 48
5.3 Project Challenges and System Limitations 49
5.4 Objective Evaluations 50

CHAPTER 6 CONCLUSION AND RECOMMENDATION 51
6.1 Conclusion 51
6.2 Recommendation and Future Work 52
6.2.1 Improvement on Hand Detection 52
6.2.2 Improvement on Object Detection 52
6.2.3 Improvement on Navigational Guidance 52

REFERENCES 53
APPENDIX A-1
WEEKLY LOG A-1
POSTER A-6
PLAGIARISM CHECK RESULT A-7
CHECK LIST A-7

LIST OF FIGURES

Figure Number	Title	Page
Figure 1.1	The Similarity Between Human Vision and Computer Vision in Processing Visual Information	4
Figure 1.2	Type of Object Detection Frameworks	5
Figure 2.1	Picture of Kinect Cane System	8
Figure 2.2	The Results of Object Recognition by Kinect Cane System	9
Figure 2.3	A Picture of Smart Cap System Prototype	11
Figure 2.4	The Overview of the Proposed Web-based Object Detection and Sound Feedback System	13
Figure 2.5	An Overview of Bbeep System	15
Figure 2.6	An Overview of the Wearable Vision-Based System	17
Figure 3.1	General System Flow Diagram	20
Figure 3.2	Use Case Diagram	22
Figure 3.3	Activity Diagram	23
Figure 4.1	Illustration of Background Subtractor Class in Forming the Background Model and Thresholding the Output Masks.	26
Figure 4.2	Code Snippet to Initialize a Neural Network	26
Figure 4.3	BlobFromImage	27
Figure 4.4	Non-Maximum Suppression	28
Figure 4.5	Morphological Opening and Closing Transformation to Produce a Clearer Mask	29
Figure 4.6	Generating a Combined Mask Using bitwise_and	30
Figure 4.7	Explanation on the Contour, Convex Hull and Convexity Defects	31
Figure 4.8	Detected Hand with Blue Points as the Detected Fingertips	31
Figure 4.9	The Direction to Reach the Origin of the Circle	33
Figure 4.10	Final Output of the System	34

Figure 5.1	Test Result for Best Brightness Condition	36
Figure 5.2	Test Result for Moderate Brightness Condition	37
Figure 5.3	Test Result for Low Brightness Condition	38
Figure 5.4	Test Result of Object Detection with Multiple Objects	45
Figure 5.5	Test Result of Object Detection with Multiple Overlapping Objects	46
Figure 5.6	Test Result of Object Detection with Multiple Overlapping Objects with Same Object Class	47

LIST OF TABLES

Table Number	Title	Page
Table 2.1	The Comparison Between the Existing Work for the Visually Impaired	19
Table 3.1	The Software Specifications for the Project Development	21
Table 3.2	The Hardware Specifications for the Project Development	21
Table 3.3	The Expected Inputs and Outcomes	24
Table 5.1	Test Results on Uniform Background	39
Table 5.2	Test Results on Messy Background	42
Table 5.3	Test Results on Different Targeted Objects	48

LIST OF SYMBOLS

θ	Theta
α	alpha

LIST OF ABBREVIATIONS

R-CNN	Region-based Convolutional Neural Network
YOLO	You Only Look Once
SSD	Single Shot Detector
RANSAC	Random Sample Consensus
Wi-Fi	Wireless Fidelity
NoIR	No Infrared
OpenCV	Open Source Computer Vision
IDE	Integrated Development Environment
API	Application Programming Interface
USB	Universal Serial Bus

CHAPTER 1 INTRODUCTION

1.1 Problem Statement and Motivation

The problem that is mostly faced by the visually impaired is most of the facilities provided to them are costly and expensive. For instance, a guide dog for blind people will eventually cost approximately $45,000 or above just for its training fees. Thus, a cost-effective solution that utilizes object detection technology is needed to help them to reduce their economic burden.

Next, a report from Zuckerman [1] also stated that 1 out of 5 blind male is living lonely and blind female is most likely to live alone as they get older. This problem is also alarming as they may find difficulties when nobody is around them to support them in doing their daily chores. Furthermore, some of the existing applications are using volunteers to help blind people through mobile devices. Indeed, object detection technology can be implemented into the mobile application to further enhance and improve the application functionality.

Besides, according to the Centers for Disease Control and Prevention [2], it is estimated that the number of adults with visual impairment and age-related eye diseases will double in the next 3 decades due to rapid ageing in the United States. This issue has triggered a necessity for a functional object finder application to be implemented to help those who have lost their vision. All of these problems are giving the motivation for an object finder system specially designed for people who are suffering from visual impairment to be deployed and operate in their daily life so that they could live more independently.

By using the object finder application, it can help the visually impaired recognize some of the daily essentials such as a cup, plate, bottle, and book and they will be guided with a computer-generated human voice to reach the objects. Furthermore, the object detection system can also potentially be implemented in a commercial industry whereby an industrial robotics arm could be embedded with an object detection module to easily navigate to a specific product and pick it up automatically.

1.2 Project Scope

The object finder application will be used to detect and identify common objects that can be found in the indoor environment like the kitchen using computer vision technology. Up to 80 classes of targeted objects are the cup, table, chair, bottle and chair in the image should be able to detected by the system. Next, the targeted objects within 2 to 5 meters should be able to be identified by the application. Also, the application should be able to identify the user's hand and compute the direction to reach the targeted object.

Finally, the output of the system ensures informative voice feedback is given to the visually impaired people in gaining a better understanding of localizing the objects in the surrounding environment, as well as providing the navigation to the users in reaching and picking up the detected objects. However, due to limited time constraints in the project development, the ability to recognise and register users' personal stuff and friends will be the future work for the project.

1.3 Project Objectives

The project aims to develop an object finder application for visually impaired people to locate and identify daily essentials efficiently in the indoor environment. The objective can be divided into 3 sub-objectives which are:

(A) To develop an object detection module that detects and identify common objects such as the table, chair and bottle in an indoor environment in the image.

- The users will capture the surrounding environment and the image captured will be the input to the system. From there, the system shall recognize the object present in the image. The detected objects will be labelled and surrounded by boxes. Common indoor objects that are located within 5 meters are desired to be detected by the system.

(B) To develop a method that track and detects the users' hand and compute the direction to reach a targeted object.

- Visually impaired people usually have difficulty in locating and positioning an object. Thus, the proposed system must recognize the user's hand that appears in the image and compute the direction from their hand to a targeted object.

(C) To develop an object finder system that provide navigational guidance and instruction based on the result to allow the users to reach the detected objects.

- When the hand of the users and the object are being detected, the system should provide the direction to reach the object through sound feedback. Meaningful and precise instruction should be provided to the users so that they can grab the object without any difficulty.

1.4 Impact, Significance and Contribution

The final deliverable in the project is a functional application that has a user-friendly human-computer interface to help people with visual impairments. This would allow the visually impaired people to live more independently and liberally without the need to employ human assistance to help them. As a result, the visually impaired will become less reliant on their caregiver or guardian to finding up an object for themselves. The system can help to identify common objects in the indoor environment, which would significantly reduce their burden in localizing and picking up the object.

At the same time, the visually impaired will be allowed to use the application without having to pay an expensive fee. The system will act as a cost-effective alternative that helps visually impaired people get rid of assistive tools that required extra charges. This would help them to reduce the economic burden that cause by accessing expensive aiding tools. Furthermore, as the system would notify the surrounding objects to the users, dangerous objects such as scissors can be avoided to prevent the users from getting hurt by the sharp objects.

The object detection module is also typically useful and can be applied in the robotics industry. The robot can easily identify the object using computer vision technology and later pick up or arrange the store item based on the requirement. This would be essential and helpful in realizing a fully-automated factory in the manufacturing or assembly industry.

1.5 Background Information

Vision is one of the most important senses for humans and other living things. Humans rely on vision to interpret and interact with the surrounding visual environment. To mimic such ability and apply it to the robots and computers, researchers and developers have focused their study on computer vision to improve robots' capability in doing their tasks. According to Brownlee [3], computer vision is a multidisciplinary field consisting of subfields of artificial intelligence which may involve the use of specialized methods and make use of general learning algorithms to process and analyse visual data (images, videos) to help the computers 'see' the world.

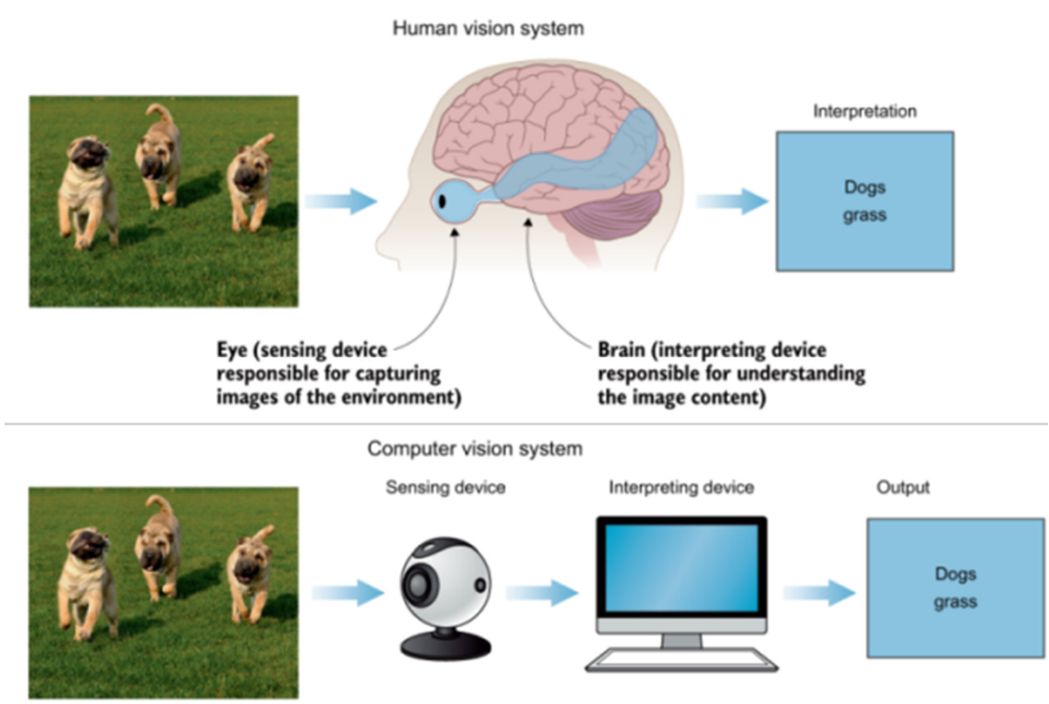


Figure 1.1 The Similarity Between Human Vision and Computer Vision in Processing Visual Information

As referred to Figure 1.1, computer vision has a similar procedure to the human vision system in interpreting visual information [4]. It aims to help machines and computers to understand the contents of digital images and videos using developed algorithms and methods. As such, one of the ways to understand the contents from the input provided is by using developed object detection algorithms to identify and locate the objects that are existed in the images and videos.

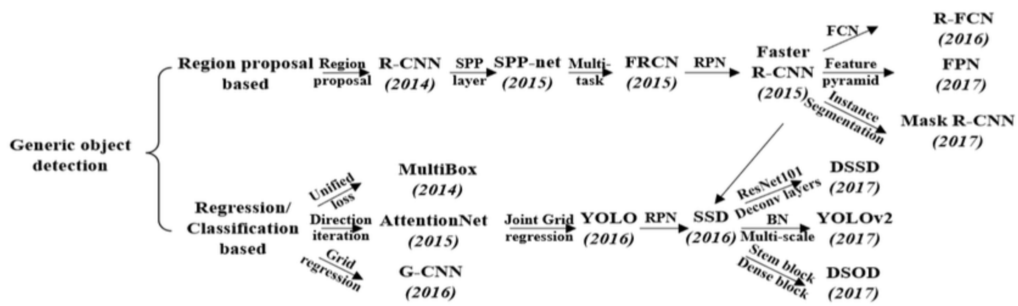


Figure 1.2 Type of Object Detection Frameworks

According to Zhao, et al. [5], object detection models are getting robust with the emergence of deep neural networks. Consequently, object detection models can learn more complex features present in different objects, allowing accurate object detection to be attainable. The unique features and characteristics owned by a specific class of objects are extracted and later help in classifying back the object itself. Object detection frameworks are separated into regional proposal based frameworks and classification based frameworks.

A regional proposal based framework proposes regions in which an object will possibly reside. Later, these regions are used to run through the classification network to detect the objects present in these regions. In addition, the Region-based Convolutional Neural Network (R-CNN) object detection algorithm is one of the illustrations of the regional proposal based framework.

Meanwhile, the classification based framework is unlike the previous framework. It directly maps the image pixels to classification and bounding box regression at the same time, making it faster and time-saving. You only look once

(YOLO) and Single Shot Detector (SSD) are the typical approach of the classification based framework.

Object detection in computer vision has a close relationship with video analysis and image understanding and thus it has a variety of uses in our daily life. It is widely applied in many industries such as video surveillance in the security industry and self-driving cars in the automotive industry. As machines can 'see' the world through real-time video and images with object detection technology, it is essential in developing a system that can help unfortunate peoples who have low vision or even lost their vision.

The object detection technology can even be integrated with a user-friendly interface into mobile applications that can rehabilitate and train users with vision impairment [6]. The system will become the alternative guide for the visually impaired to interpret and identify the objects in their surrounding environment. With a proper way to implement it, it will bring enormous benefits to the visually impaired in doing their daily chores. For example, reaching to the seat, picking up bottles from the table or locating the position of a tool.

In realizing the object detection functionality in the proposed system, the YOLO algorithm will be employed in perform object detection from the images and videos of the visually impaired or blind people. According to Redmon et al. [7], YOLO is a fast and simple object detection algorithm that identifies the objects in videos and images and creates a boundary to surround the objects. This is essentially useful for the visually impaired to find the surrounding objects quickly and accurately. When it is employed in robots or machines, it gives the ability for the robots to recognize objects. With this, an extensive functionality can be further implemented such as allowing robotic arms to locate and pick up the objects for humans.

Blind peoples are similar to robots without vision. If they could get a clear vision of the surrounding, they can do their job more professionally and efficiently. Thus, the YOLO algorithm is particularly useful in realizing the object finder application for the visually impaired.

1.6 Report Organization

In this paper, Chapter 1 starts the report by giving a brief introduction to the reader on the problem faced by people who have visual impairments as well as how computer vision could take a part in helping the unfortunate group of people. The following Chapter 2 discusses in detail the effort done by other researchers in creating an assistive system to help the visually impaired. A comparative summary and discussion are done to compare each of the existing systems. In Chapter 3, the author proposed the methodology and approach used to develop the project. The system design and operation are discussed later in Chapter 4 which the implementation and operation are shown to archive subfunction of the system module. After that, testing and performance evaluation of the system is presented in Chapter 5 by carrying on different test cases to proofread the system functionality. Finally, Chapter 6 concludes this paper and discusses the future implementation that could be done.

CHAPTER 2 LITERATURE REVIEW

2.1 Review on Previous Works

2.1.1 Kinect Cane System [8]

The Kinect Cane System [8] is a proposed assistive system that helps visually impaired people identify obstacles nearby. The system collects input from the infrared-based range sensor which is known as the Kinect sensor and it is attached to the white cane (a visual aid long cane for blind people) to help the vision loss people to avoid obstacles. The input from the sensor will be sent to the portable computer via a wired connection to process the input.

Besides detecting obstacles from the environment, the system also has a keypad for the users to recognize some objects such as seats by pressing the keypad to find the specific objects. In a normal operation, the Kinect sensor will input the scene image and depth data and sent it to the computer for algorithm computation. If the system detects the obstacle, the users will get feedback through vibration from the tactile device. Also, if the users know there is a seat nearby but could not locate it, they can press the corresponding key on the keypad to recognize and locate it.

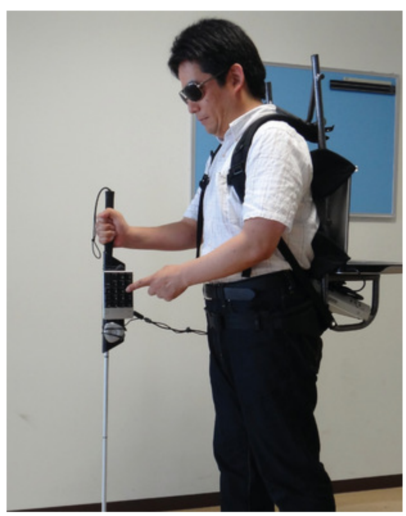


Figure 2.1 Picture of Kinect Cane System

In the object detection implementation, the Kinect sensor is used to obtain the depth data. The depth data are coded into coloured pixel which is known as the 'edges' that represent the distances from the sensor. This allows the system to determine the plane objects (such as floors and walls) by using the depth data obtained from the scene. The system then performs a series of methods based on random sample consensus (RANSAC) to recognize the planes. Once the planes have been recognized, the recognition result of planes will remain with some black pixeled edges which are used for seat recognition. The remaining edges in the region that fulfil 30 to 50 cm from the floor, and the region more than 1200cm^2 will be painted in red pixels, which is recognized as the sitting surface of the seat.

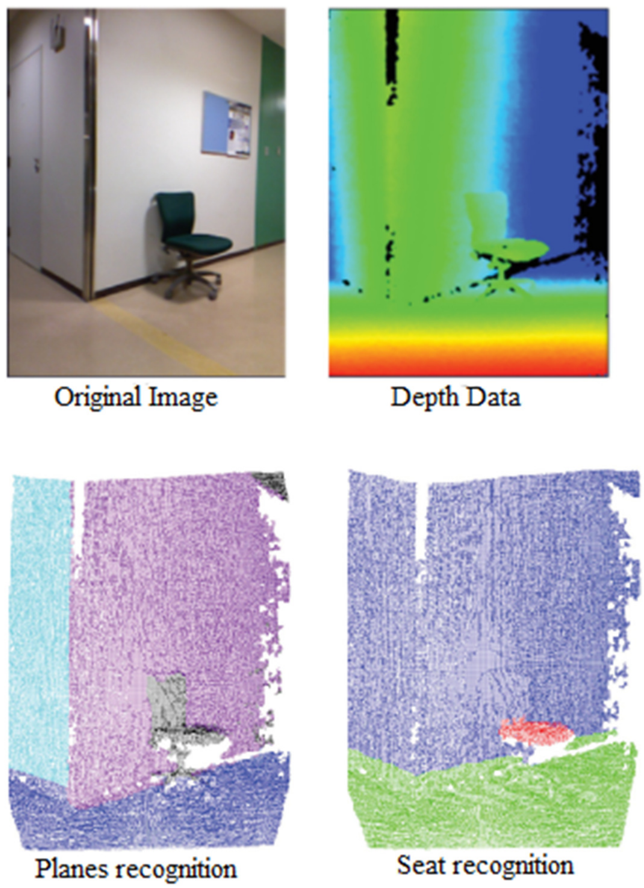


Figure 2.2 The Results of Object Recognition by Kinect Cane System

The strength of the Kinect Cane system is it has 2 detection algorithms that process simultaneously to detect the obstacle in different considerations. This would allow the system to identify large and small obstacles that can block the person's path. This is helpful for the users to avoid concealing obstacles that they could not identify by themselves when using a normal white cane. Besides that, the system can detect and recognize some essential public facilities for the users like seats, upward or downward staircase and the elevator. With this, the visually impaired can easily reach the objects and access the facilities. Then, the next advantage of the Kinect Cane system is it is a system that is extended with the use of a white cane which is commonly used by people with visual impairment, hence they can easily adapt and familiarize themselves with the system in a short period of time.

However, the weakness that existed in the Kinect Cane system is the mobility of the system. The visually impaired will need to carry a backpack to hold the portable computer and battery while walking which is inconvenient and troublesome. Next, the system passively starts to recognize the objects only when the user input from the keypad. This is a disadvantage in the system whereby if the user is not familiar with the surrounding environment, they might not have the prior knowledge of what objects or facilities is located nearby, thus they could not instruct the system to find a specific item. Lastly, the wired connection between the sensor and the computer will limit the detection range of the white cane. As a result, the white cane will be restricted to a certain area to detect and recognize objects.

An improvement can be done to the system is, if the personal computer used in the system can be replaced or implemented in a lighter and portable device such as a mobile phone or remote processing computer to execute the program and computation. Secondly, the system should actively recognize nearby objects instead of the users must instruct it to do so. The system can pre-train with more datasets so that it can recognize and detect more objects for the visually impaired. Finally, the system can establish wireless connections such as Wi-Fi or Bluetooth connection with the sensor to resolve the problem related to mobility and connectivity.

2.1.2 Smart Cap System [9]

To help the sightless and visually impaired people, Nishajith et al. [9] have come out with the idea of Smart Cap which is equipment for the blind and a guidance system for them. The system will help the blind to navigate using real-time object detection and provide voice information to the users to identify the object.

It is built with a Raspberry Pi-3 processor and connected with a NoIR camera that captures the input from the real-time environment. Then, the system is attached to a cap for the people to wear and walk around. The real-time video captured from the surrounding environment will be converted to a set of frames and then the processor will process the image and bound the objects with boxes. Finally, the identified objects will be switched to text description and provides the voice output to the users by conducting text to voice conversion. The users will be hearing the audio output by using an earphone.

Figure 2.3 A Picture of Smart Cap System Prototype

In the proposed system, the TensorFlow framework is used to perform object detection and recognition. It has offered a dataset which is the ssd_mobilenet_v1_coco model that is used to detect the object present in the input images by using a detection graph and weight. Therefore, the single shot detector (SSD) algorithm was used to implement the object detection with the help of OpenCV. This helps object detection to be done faster as compared to the traditional detection method. Next, the detected objects are stored in text format and then converted into auditory output using eSpeak, which is a speech synthesizer software.

The strength of the proposed system is it can identify about 90 classes of common objects like vases and mobile phones for sightless people to help in avoiding or picking up daily essentials. This is because the system is pre-trained using the Common Objects in Context (COCO) dataset and the number of objects can be increased by doing training of the system in the future. Besides that, the system requires no preliminary knowledge of the technology because the system will start operating when power is on. This greatly improves the ease of usage for blind people in using the guidance system. Furthermore, the whole design of the system is cheap and small which has also met the visual aid requirement for the visually impaired people in terms of mobility and affordability.

In contrast, the problem with the power supply will be the limitation that exists in the system. The power consumption of the real-time camera to constantly captures video from the scene surrounding is high however the power supply is only provided by using a power bank. The power supply will not able to fulfil the system requirement if it needs to operate for a long time. Next, there is still a reliability issue with the system whereby the creator of the system has mentioned that there is a flaw and misinterpretation of the system in identifying an object occasionally. This problem will lead to misinformation and wrong interpretation done by the vision loss people which may bring harm to them.

To overcome the issues that have been mentioned above, the system can firstly adopt a power supply that has longer life spend or replace it with a renewable power source such as a solar energy panel that captures sunlight to supply the power usage. This would make the system to be feasible for a long-time operation. In addition, the system should also do more training with more datasets to make assurance for the system in terms of the system's reliability and safety.

2.1.3 Web-based Object Detection and Sound Feedback System [10]

Intending to convert the visual environment to an auditory environment for the visually impaired people to understand their surroundings, Hoa et al. [10] have proposed a low cost and flexible system in mobile devices that recognize the objects and provide sound feedback to support the blind users.

The users can use the system by holding the mobile smartphone in front of their chest, with an application that records real-time video from the smartphone camera. The video will then be transmitted to the server and the objects contained in the video will be identified by using a server-sided object detection model. From the server itself, the detected objects will be matched with the corresponding sound file and finally feedback to the user through the smartphone speaker.

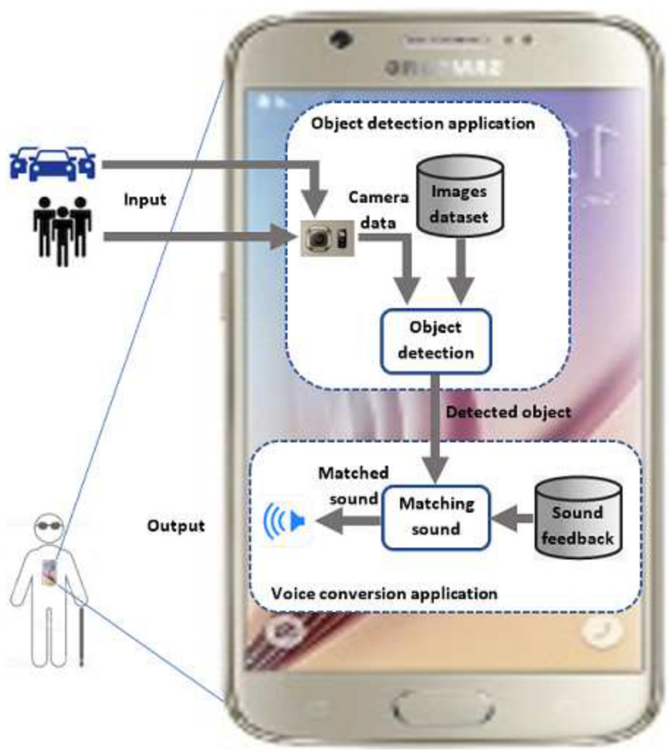


Figure 2.4 The Overview of the Proposed Web-based Object Detection and Sound Feedback System

Since the object detection model is implemented in a web environment, the author used TensorFlow JS which is a WebGL (Web Graphics Library) that enables the development of image processing projects. Subsequently, the author used SSD-MobileNetV2 to implement web-based object detection. To allow input video from the client's mobile phone to be transmitted to the server, Ngrok (an application that allows establishing of local web service) is used and the author deploys the object detection system to the Ngrok network to process the input. The server will then detect the object and sent the corresponding sound back to the users.

The strength of the proposed system is its flexibility and mobility of the system. The users can directly install the application to any smartphone and used it when needed. This is also convenient for the users to use the system as the smartphone are lightweight and easy to carry while walking outdoors. As the processing is done on the server side, the client devices are free from lag and insufficient processing power. This would ensure a fast processing speed. Another advantage of the system is it classifies the same type of objects and then feedback on the quantity of the object accordingly. This avoids confusion and chaotic information being delivered to the users so they can know the exact number of that particular object which exists at the moment.

The disadvantage in using the system is the issue that deals with the internet connectivity. The application will eventually become useless as all of the video and sound implementations were located on the server. The system is unable to operate in a poor internet connection or airplane mode. Moreover, the number of objects existing in the video will drastically affect the accuracy of the system in identifying the object. If the number of objects increases, there will be an increase in the number of undetected objects too.

To resolve the problem, the system should consider adding the object detection model in the mobile application as well hence the system can still perform without an internet connection. With this, the issue of internet connectivity will not affect the normal use of the system for the vision loss people in navigating themselves independently. Then, the result from the mobile phone can later be sent to the server to perform further computation and benchmarking the improve the accuracy and reliability of the system.

2.1.4 Assistive Suitcase System, Bbeep [11]

Bbeep [11] is a vision-sensing system to track the movement and motion of pedestrians and predicts their possible future position and later generates an alarm to warn the people and clear the path in front of the blind user to allow them to walk through a crowded environment. The system uses a stereo camera that is attached to the suitcase to capture the real-time data. From there, the system can detect the pedestrians and their position to predict their future positions based on the result from the camera.

Then, the system determines the risk of collision with the blind user and pedestrians and emits the audible signal based on the situation. The alarming sound is categorised into 3 types which are a low-urgency beep, intermediate-urgency beep and stop sound. Different sounds will be emitted based on different scenarios, for example, if there is a chance of collision between the blind user and the pedestrians within 5 seconds, low-urgency beep will be emitted. The sound will change according to the time interval between the collision respectively.

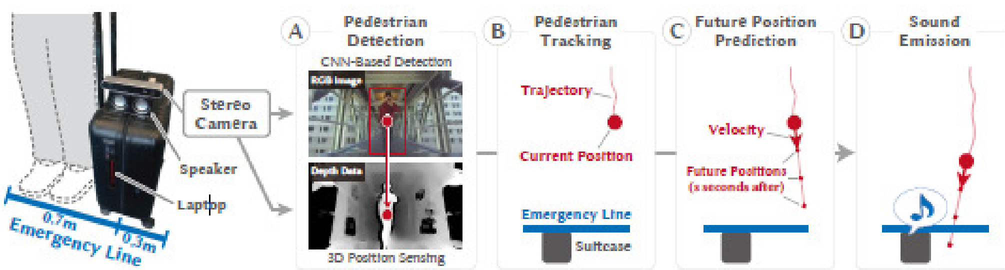


Figure 2.5 An Overview of Bbeep System

The system has installed a stereo camera that supports 3D odometry to trace the movement of the camera in real-time. This eliminates the effect of suitcase rotation. YOLOv2 is used in the system to detect the incoming pedestrians and obtain their 3D position in the camera coordinate system. The system constantly updates the 3D positions of the pedestrians when the detection result has been updated, thus allowing the system to detect and track the position of pedestrians simultaneously.

The pros of the system are it can predict the risk of collision with the pedestrians and obstacles such as chairs and walls, and make a clean and safer path for the visually impaired users to navigate through. The warning sound is not only giving awareness to the user itself, however, other sight pedestrians can also be alert and notice the user, hence both parties can take necessary action in navigating through a crowded place. Next, the situational alarming sound will be provided to the users so that they can have a better understanding of the surrounding environment.

On the other hand, the system has some limitation that must be resolved to further improves its performance of the system. Firstly, the system can be better if it can include what objects have been detected and feedback to the users. This is significant because the users might not exactly know the situation with only the alarming sound provided by the system. Secondly, this system is not suitable to be used in some quiet environments such as hospitals and libraries. This is because the alarm sound will notify the users and the nearby pedestrians, which may directly disturb them in a quiet place. The alarm sound will also attract too much attention from the others which might embarrass and burden the users and other people. Lastly, the volume of the suitcase and laptop is big and they are not suitable and inconvenience for the user to carry and walking freely.

The suggestion to make the system better is the system can provide informative feedback on the detected objects to the users to let them gain a better understanding of the surrounding environment. Besides that, the system can output the sound to a speaker or headsets which has lower volume so that it would not be too disturbing to the users and the pedestrians. Also, the volume of the alarm sound should be allowed to be adjusted by the users if necessary. It is also significant to make possible integration of the system with other mobile devices such as the smartphone to further improve mobility. Therefore, the system can be used anywhere and on different platform which is suitable for the users to walk liberally.

2.1.5 Wearable Vision-based System [12]

Wang et al. [12] have come out with a wearable system that is implemented using computer vision technology and haptic feedback to the blind users to help them in identifying free space, avoid obstacle collisions in trajectory and locate certain types of objects. The whole system is implemented by getting input from the depth camera that is held on the user's chest, after passing through the system algorithm, the system output the results (Free space parsing or the object detected) through the haptic vibration bell and braille display (tactile writing system for the visually impaired) that wore around the user abdomen.

The haptic vibration bell will vibrate from the left, middle and right which corresponds to the obstacle's position from the sagittal plane of the user at -30°, 0°, 30° respectively. When there is no vibration, it indicates that there is free space around the user. Besides that, the user will get feedback from the braille display of the system to identify an object. The braille display is encoded with 4 different types of objects which are represented with different braille symbols.

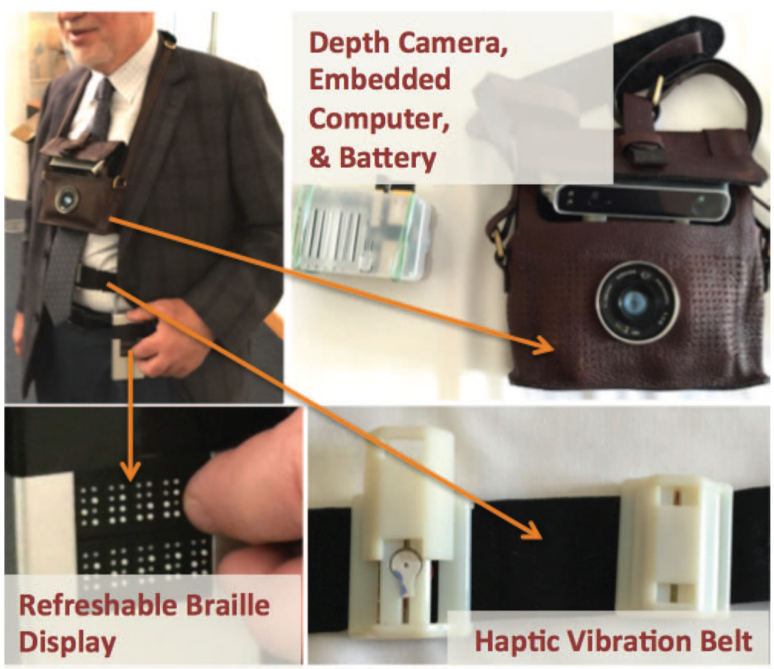


Figure 2.6 An Overview of the Wearable Vision-Based System

The strength of this system is it can accurately navigate the blind to free and walkable space to avoid collisions with obstacles. The users can adapt easily to the use of the system. Moreover, this system is not only limited to straight-line navigation however it is robust enough to navigate the user on a complex path with many turns. This allows the user to have safe navigation of making turns and even walking in a maze. Furthermore, the haptic device has low latency and fast response, providing up to time information to the users in avoiding the obstacles and directing them to the free space.

A limitation that existed and needs to be addressed in the system is it is only restricted to the indoor environment. The system is unable to perform well in an open environment due to exposure to sunlight will reduce the accuracy of the camera. Unclear and blur images will also be captured due to the body movement of the user. Also, the view of the camera is restricted and hanging obstacles above the chest will be undetected. Adding on, there is only a limited number of objects that can be notified to the user through the braille display. Hence, the user might not get a better understanding of the object in the surrounding environment.

The improvement that should be done in the system is firstly the camera should be upgraded so that it can support high real-time video input no matter in the indoor or outdoor environment. Next, the camera can position on a higher part of the body such as the head to eliminate blind spots on upper part of the body while capturing input. The system can also include informative audio output to the users as an alternative way to further notify and provide useful information to the users while travelling.

2.2 Table of Comparison Between the Existing Works

To compare each of the proposed systems, the table of comparison is formed as shown below. Each of the existing systems will be evaluated based on 5 criteria which are the operation mode, object detection coverage, the type of detected object, its mobility and the type of feedback provided.

Table 2.1 The Comparison Between the Existing Work for the Visually Impaired

System	Operation mode	Coverage	Type of detected object	Mobility	Type of feedback
Kinect Cane System [8]	Offline	Indoor and Outdoor	Stationary	Low	Vibration feedback
Smart Cap System [9]	Offline	Indoor and Outdoor	Stationary	High	Sound feedback
Web-based Object Detection and Sound Feedback System [10]	Online	Outdoor	Stationary and moving objects	High	Sound feedback
Bbeep System [11]	Offline	Outdoor	Stationary and moving objects	Low	Sound feedback
Wearable Vision-Based System [12]	Offline	Indoor	Stationary	High	Vibration & Haptic feedback

From the table above, the first requirement of the assistive system is the system should be able to operate in offline mode to eliminate the reliability of the system on the internet connectivity. Secondly, the system must be able to operate in both indoor and outdoor environments so that the users can use the system anywhere they are travelling. Detecting both unmoving and moving objects are one of the key components of the system to provide more useful information to the users. Since the users will be carrying the system while travelling, the system should integrate with high mobility and provides sound feedback for the user regarding the surrounding environment.

CHAPTER 3 SYSTEM METHODOLOGY/APPROACH

3.1 Design Specifications

3.1.1 Methodologies and General Work Procedures

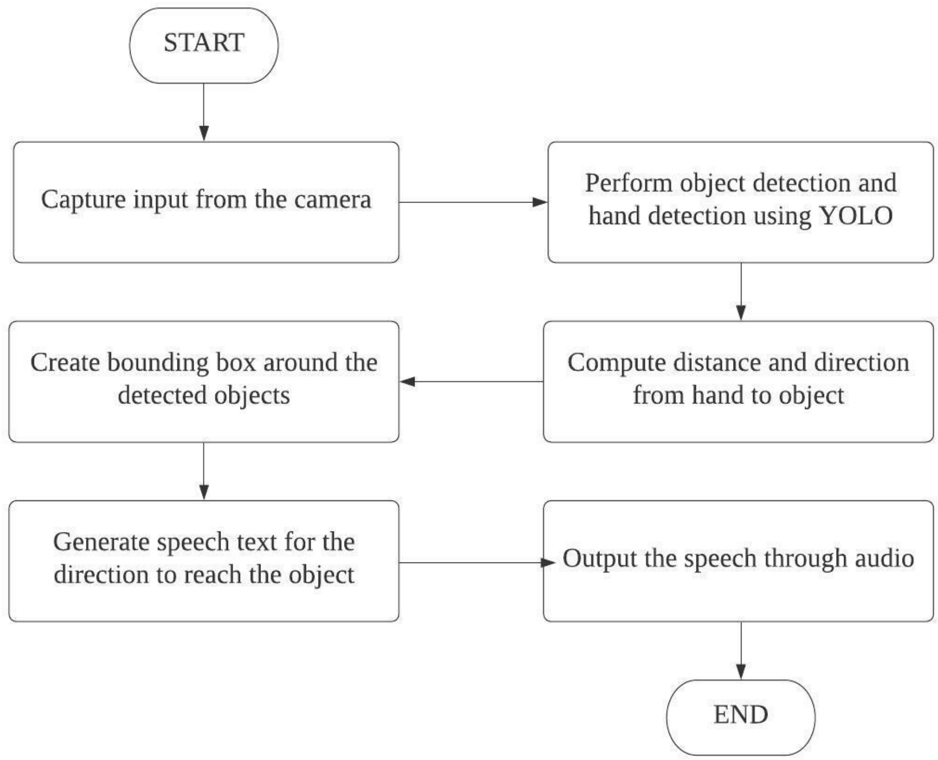


Figure 3.1 General System Flow Diagram

The system starts by capturing the input of a set of images or video from the camera. The input will be pre-processed before passing through the YOLO network. After this, the input will be passed through a pre-trained YOLO network to detect the objects or the user's hand present in the input. If the user's hand is detected in the image, the system will compute the shortest path and direction to reach the object.

Meanwhile, the detected object will also be bounded with square boxes in the image. The detected object and its direction from the hand will then be translated into speech text. For example, 'The spoon (detected object) is on the left (direction)'. Finally, the system will speak out the speech to the users to help them in navigating to the object.

3.1.2 Tools to Use

Table 3.1 The Software Specifications for the Project Development

Tools	Description
Visual Studio 2019	Throughout the project, Microsoft Visual Studio 2019 will be used as the integrated development environment (IDE) to develop the system in C++. The source code implementation and debugging process will be done using this IDE. Visual Studio Community edition will be used which is free of charge.
OpenCV	OpenCV is an open source library which helps to archive real-time computer vision implementation. It will be used in image processing and object detection implementation in the proposed system.
You Only Look Once (YOLO)	YOLO is an object detection method that is used to detects and recognize objects from the visual data. It is a pre-trained model that is fast and can recognize up to 80 classes of different objects which are commonly found in the real world.

Table 3.2 The Hardware Specifications for the Project Development

Hardware Specifications (Laptop)	
Model Name	Acer Aspire E
Processor	Intel i5-7200U 2.5GHz
Memory	12GB
Storage	128 GB SSD + 1TB HDD
Operating System	Window 10 64 bits

3.1.3 User Requirements

- The users must be able to provide input to the system in either videos or images.
- The system must be able to identify objects that are found within 2 to 5 meters in the image.
- The system should be able to recognize the hand of the user in the image
- The system should determine the direction of the object from the user's hand.
- The users must receive an auditory output from the system regarding the image information.
- The system should output errors to the users if objects were not detected.

3.2 System Design Diagram

3.2.1 Use Case Diagram

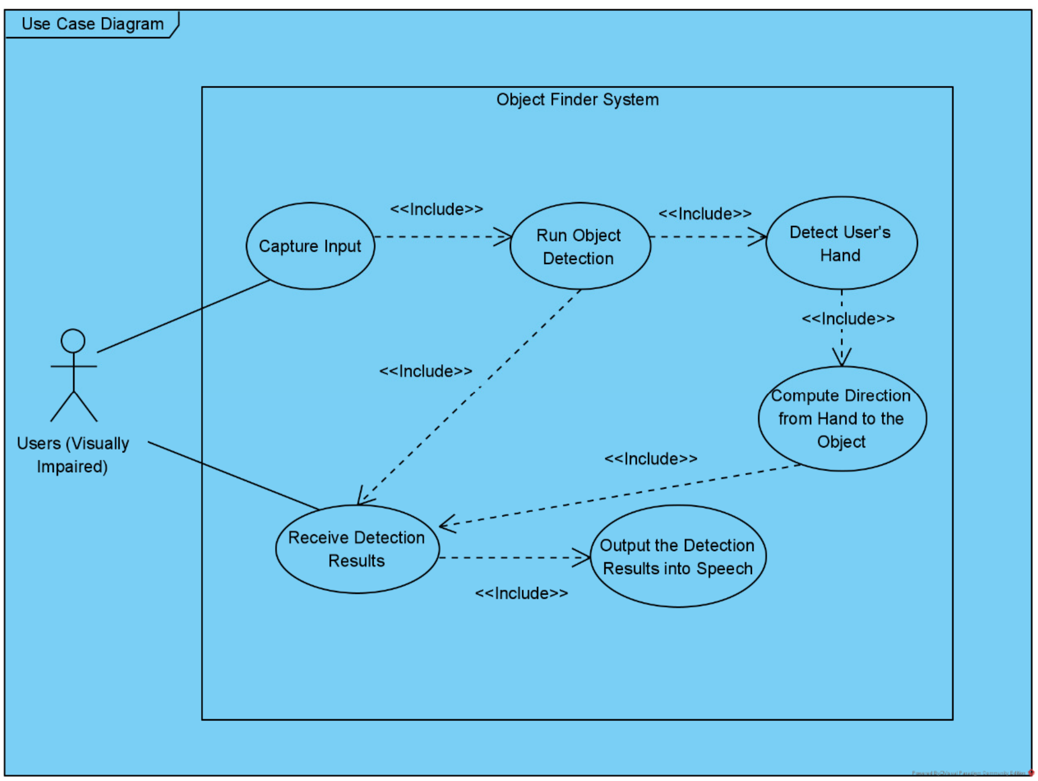


Figure 3.2 Use Case Diagram

The use case diagram above summarizes the functionalities of the object finder system. It is consisting of 5 use cases which are capture image, run object detection, run hand detection, compute direction from hand to object, receive detection results and read out detection results.

3.2.2 Activity Diagram

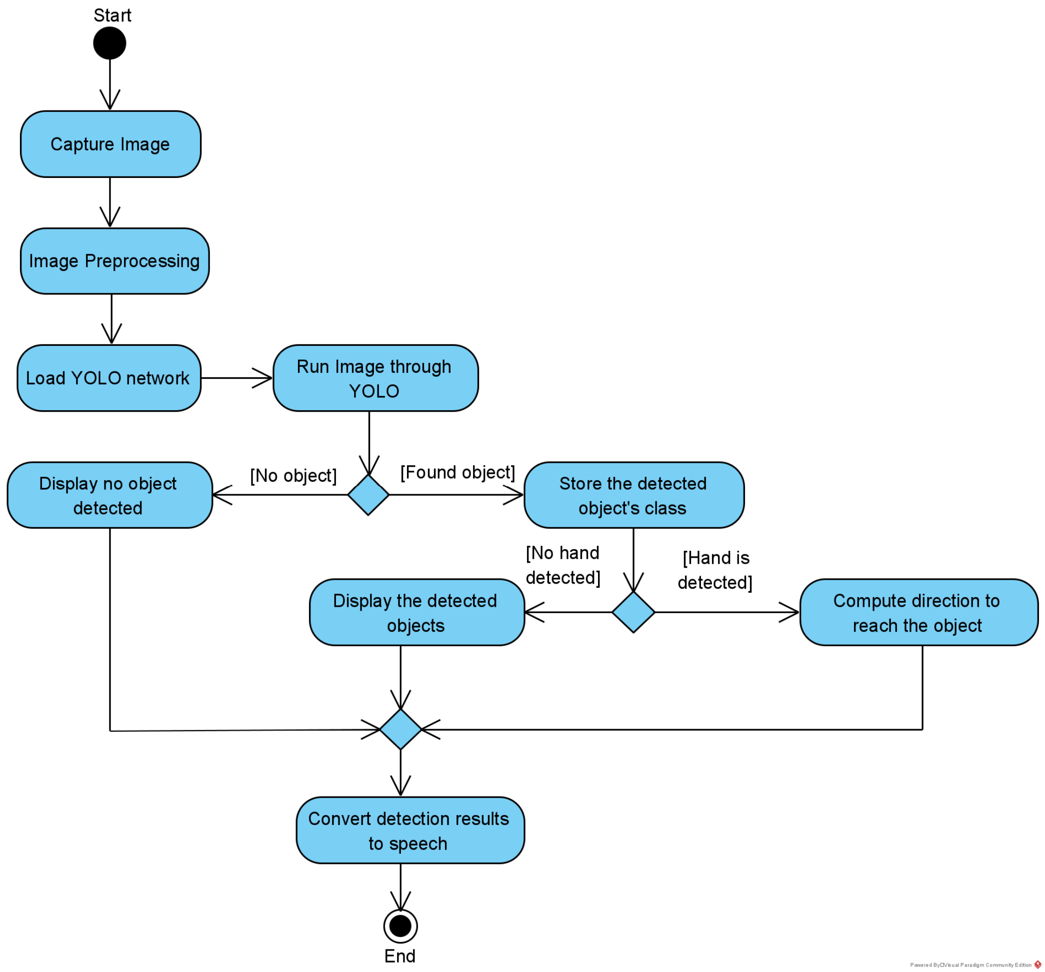


Figure 3.3 Activity Diagram

The activity diagram above generalizes the flow of the system operation dealing with different inputs and conditions. For example, if the object detection model could not detect any available object nearby, it will output the detection result by speech. On the contrary, the system will compute the direction of the user's hand to reach the object. Finally, all of the detection results will be stored as a text and it will be convert into speech output to inform the user.

3.3 Verification Plan

Table 3.3 The Expected Inputs and Outcomes

Inputs	Expected Outcomes
The system receives input frame with detectable objects.	The system should detect the object and surround it with square boxes.
The system receives input frame without detectable objects.	The system should display with no result found.
The system receives input frame with multiple detectable objects	The system should detect all of the object presented and allow the user to choose only one target object
User hands are found in the image with an object nearby.	The system should provide the direction from the user hand to reach the object.
User hands are overlapped with the object in the image	The system should notify the user so they could grab the object successfully.

CHAPTER 4 SYSTEM IMPLEMENTATION

4.1 Overview

In order to archive the objectives of the object finder system, some assumptions are made to minimize the implementation challenge. These assumptions are stated as below:

- The camera is remained static and does not move during the process.
- The detected object is stationery and restrained in the same position in each camera capture.
- The brightness and illumination from the surrounding environment do not change extremely over time.

With this, the system is separated into different modules, which are the background subtraction, object detection, hand detection and compute the direction.

4.2 Background Subtraction

Since the camera is assumed to be static, the system can detect the user's hand by subtracting the background from the foreground. To do so, the system needs to receive inputs of the background frame captured from the USB camera to learn the background model. This will help the system to detect the hand motion from the user later as a foreground.

BackgroundSubstractorMOG2 class provided by OpenCV is declared to perform the background subtraction task which takes in 60 frames of the image to compute the background model and with a threshold value of 16. It detects whether each pixel from the input frame is from the background or the foreground one by computing the Mahalanobis distance between the pixel and the background model. Then, input images are captured to apply to the background model with a learning rate of -1 which make the background algorithm choose the learning rate automatically. The background model will be computed and it will be used for object detection to determine the objects that are presented in the background.

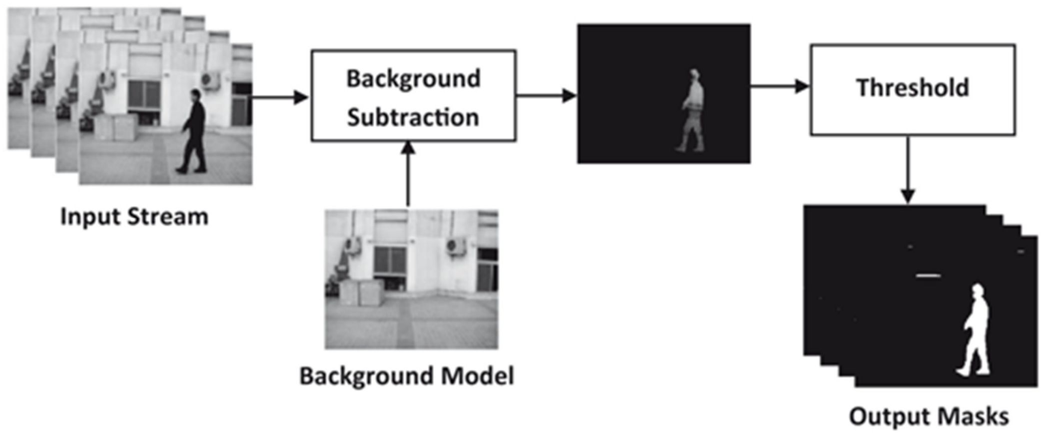


Figure 4.1 Illustration of Background Subtractor Class in Forming the Background Model and Thresholding the Output Masks.

4.3 Object Detection

The system used a pre-trained YOLO model to perform object detection using a neural network. Therefore, the system needs to set up a neural network to run the object detection model. The pre-trained YOLO model weights and configuration files must be loaded to initialize the network. The weights and configuration files are available at https://pjreddie.com/darknet/yolo/.

4.3.1 Setup Neural Network for YOLO Model

The neural network is created using the *readNet* function to read the trained weights and configuration files. The network backend and target that are used for computation are also specified using *setPreferableBackend* and *setPreferableTarget*. The code snippet below shows the way of setting up the neural network.

```
String weights = "yolov3.weights";
String configuration = "yolov3.cfg";

cv::dnn::Net net = cv::dnn::readNet(weights, configuration);
net.setPreferableBackend(cv::dnn::DNN_BACKEND_OPENCV);
net.setPreferableTarget(cv::dnn::DNN_TARGET_CPU);
```

Figure 4.2 Code Snippet to Initialize a Neural Network

4.3.2 OpenCV blobFromImage Preprocessing

The input frame that will be used to detect objects is obtained from the background model to determine the objects that exist in the background. The input frame needs to perform scaling, mean subtraction and channel swap using the *blobFromImage* function provided by OpenCV. The scale is set to 1 / 255 to normalize the intensity values of the input frame into the range of values between 0 to 1. Since the YOLO model is using the input layer with a scale of 608 x 608, thus, the weight and height of Size(608, 608) are set in the input layer to process the input frame into a blob 4D array. The mean subtraction can be ignored. The swapRB is set to true to produce BGR to RGB transformation in the colour channel because the OpenCV sometimes uses BGR colour schema instead of RGB. Finally, the blob is set to the neural network using *setInput* to evaluate the input frame.

```
bg->getBackgroundImage(frame);
bool swapRB = true;
blob = cv::dnn::blobFromImage(frame, 1 / 255.0, cv::Size(608, 608), Scalar(0, 0, 0), swapRB, false);
net.setInput(blob);
```

Figure 4.3 BlobFromImage

4.3.3 Evaluate Detection Results

The image blobs are forward into the YOLO network to perform detection. YOLO model divides the image into many regions then predict the bounding boxes and probabilities and return the output as shown as below:

$$[x, y, w, h, class1, class2, \dots, classN]$$

where

x is the x coordinate of the bounding box

y is the y-coordinate of the bounding box

w is the width of the bounding box

h is the height of the bounding box

class1, class2, …, classN is the confidence score from 1 to N object classes

This indicates that the object with a higher confidence score would consider existing in the image. MinMaxLoc function from OpenCV is employed to retrieve the maximum confidence score in each of the proposed regions. The system will then save the class of the object which has a confidence score higher than 0.5. Before displaying the final result to the users, the detection results need to be passed into *NMSBoxes* function to perform non-maximum suppression with a threshold of 0.4 to remove redundant and overlapping detection results. Finally, the detected objects are surrounded by square boxes with their corresponding confidence score.

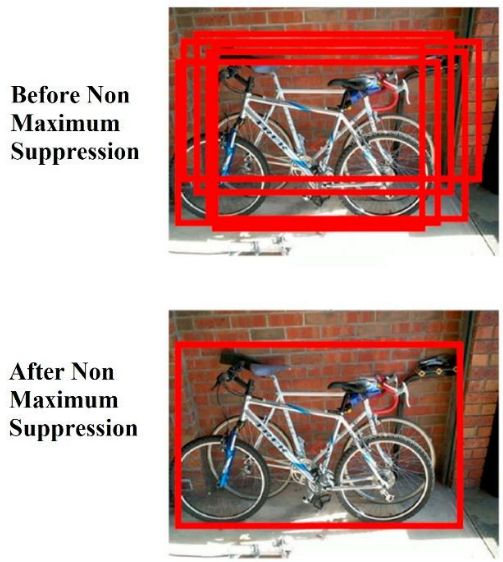


Figure 4.4 Non-Maximum Suppression

4.4 Hand detection

Once the objects are identified and their positions are determined, the system will proceed to find motion from the camera captures and detect the user's hand from there. To differentiate between the user's hand and the background pixels, the background subtractor is applied to produce a foreground mask but the result will possibly affect by the change of brightness and shadows. As a result, image segmentation is performed to further improve accuracy in detecting the hand region.

4.4.1 Skin Colour Detection using YCbCr Colour Space

The system is required to perform skin colour detection using image segmentation to allow the region of user's hand motion and movement to be correctly located. Apart from all of the available colour space, the YCbCr colour space is chosen to perform skin segmentation as it has outperformed than the other colour space [13]. The proposed range of values for YCbCr colour space that best represent human skin colour are shown as the following:

$$16 \leq Y \leq 235$$
$$77 \leq Cb \leq 127$$
$$133 \leq Cr \leq 173$$

The system first converts the camera feed into YCbCr colour space and perform gaussian blur to smoothen each of the image frame. Then, thresholding is performed by calling *inRange* function from OpenCV and pass in these values to generate a skin mask which represents the skin colour region. The generated mask will undergo morphological transformation to remove noise and outputs a clean mask. Later, the skin mask is combined with the foreground mask from the background subtractor using *bitwise_and* function to produce the final binary mask that represents the hand region.

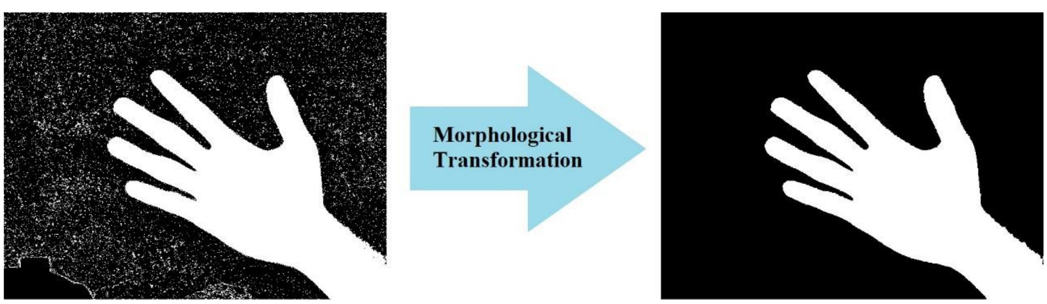


Figure 4.5 Morphological Opening and Closing Transformation to Produce a Clearer Mask

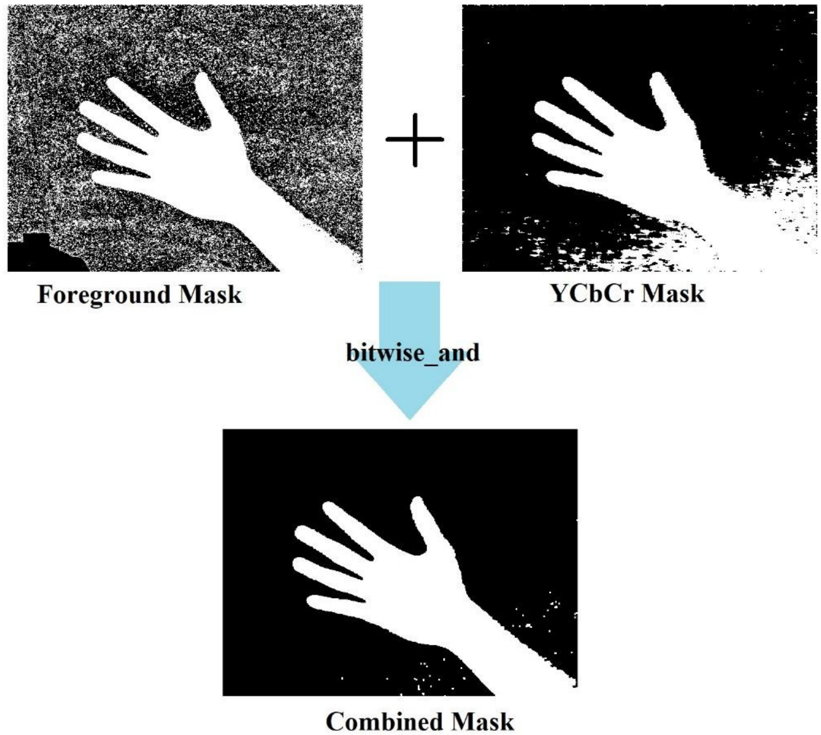


Figure 4.6 Generating a Combined Mask Using bitwise_and

4.4.2 Hand Contour and Finger Detection

The system can now identify the user's hand region more precisely and accurately. Therefore, the available contours from the binary mask is extracted using *findContours* function to collects all of the contours exist in the binary mask. The system calculates each of the contour's area and return the contour with the largest area. Assumption is made that the hand contour will occupy the largest contour in the binary mask.

To obtain the fingertip location, the system needs to perform *convexHull* to find the minimum enclosing boundary that warp out the hand region. If so, the finger point should have lie within the convex hull boundaries and obtaining the fingertips point is possible by performing *convexityDefects* function provided by OpenCV library. The function returns the start point, end point, farthest point, approximate distance to farthest point of each deviation in the convex hull. The starting point and ending point are used to calculate the angle in between, θ and the point will then be identified as the fingertip if such that θ is an acute angle. The final results obtained are quite satisfactory and precise.

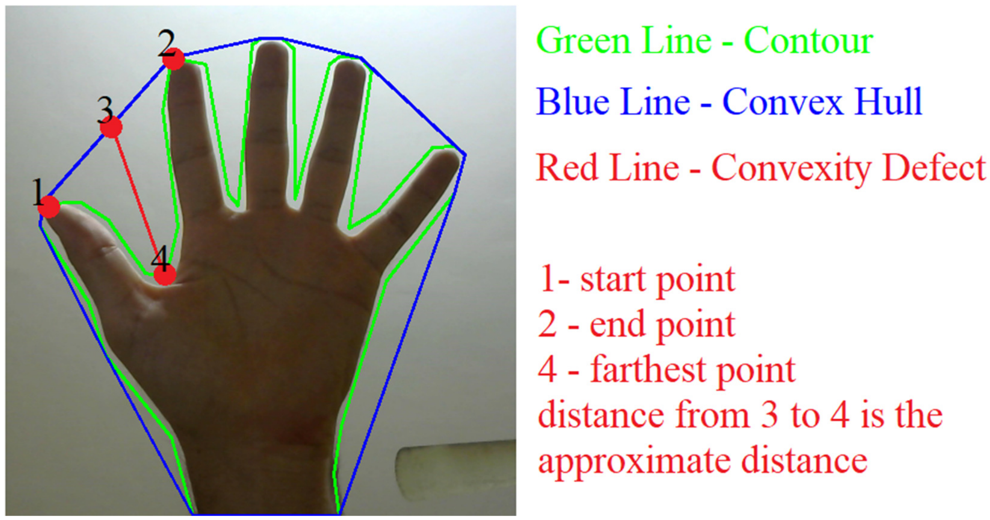


Figure 4.7 Explanation on the Contour, Convex Hull and Convexity Defects

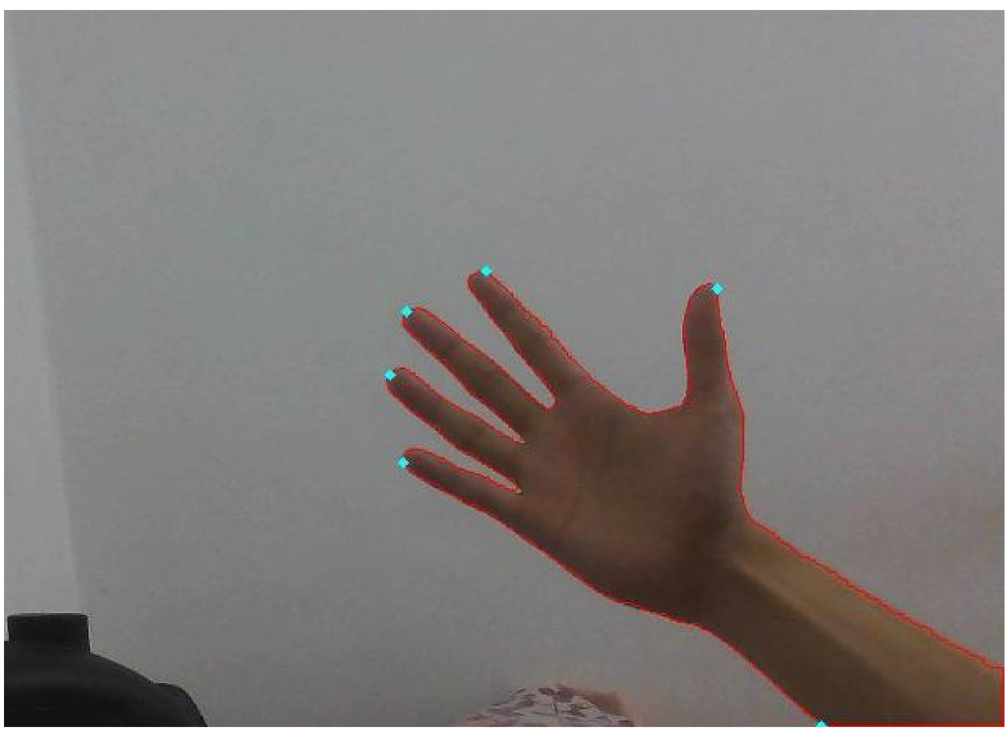


Figure 4.8 Detected Hand with Blue Points as the Detected Fingertips

4.5 Compute Direction

Once the system has successfully located the user's hand and the objects, the system can simply compute the direction from the hand to reach the objects. To make the system more feasible, only 1 object will be selected as the targeted object to compute its direction from the hand and the user is able to switch between different target object by pressing space button. To avoid the scenario where the system fails to identify the fingertip location, the system will calculate the centroid point of the hand contour and compute its direction to reach the object. The center point of the hand contour can be calculating with the image moments where:

$$Cx = \frac{M10}{M00}$$

$$Cy = \frac{M01}{M00}$$

where

C_x is the x-coordinates of the center point

C_y is the y-coordinates of the center point

M denotes the image Moment

Once the center point of the hand is obtained, the system then computes the angle between the hand's center point and the targeted object center point. As the image frame has loses 3D spatial information, the direction is limited into 2D directions which are up, down, left and right. The relative direction of the hand and object is determined as shown as the image below. The direction indicates where the user's hand should move and the origin indicates where the object is positioned.

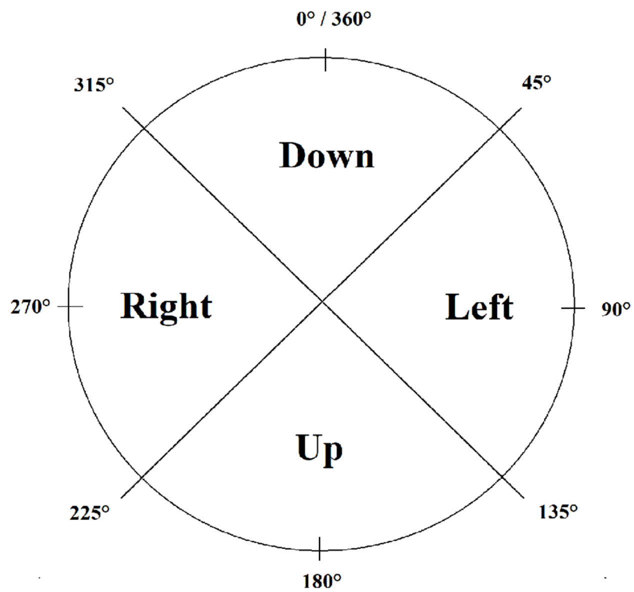


Figure 4.9 The Direction to Reach the Origin of the Circle

Let α to be the angle between the hand and the object, if α is:

315° < α < 360° OR 0° < α < 45°, then the system will issue "Go down".

45° ≤ α ≤ 135°, then the system will issue "Go left".

135° < α < 225°, then the system will issue "Go up".

225° ≤ α ≤ 315°, then the system will issue "Go right".

Finally, the system will draw a connected line from the hand to the object. Voice output will also be generated asynchronously to tell the users to move their hand in a specified direction to be able to grab the object.

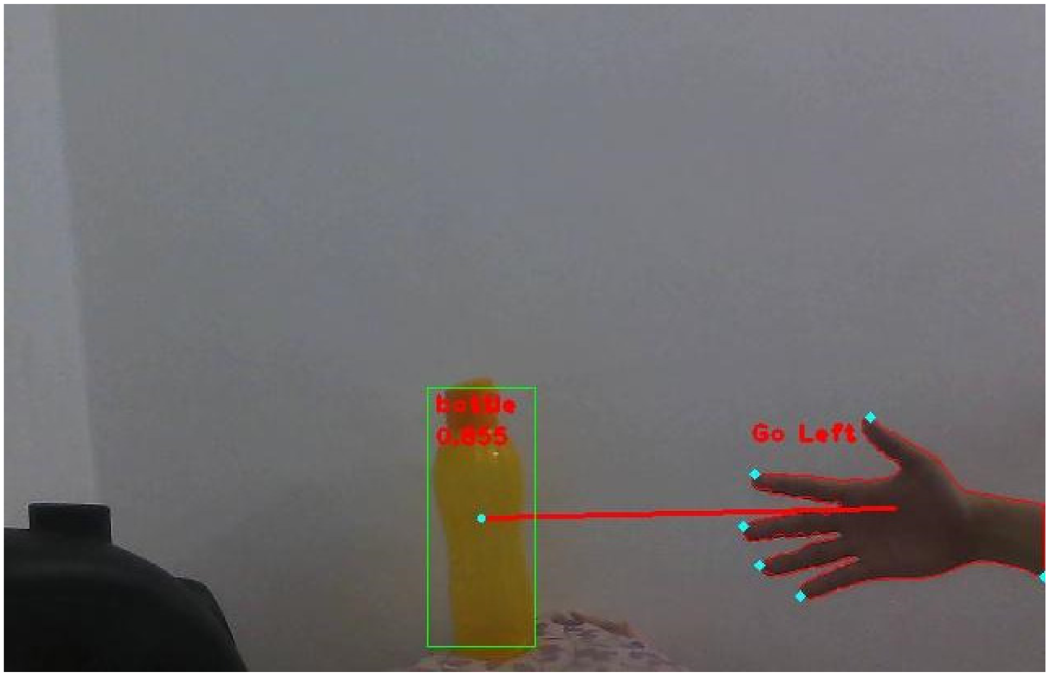


Figure 4.10 Final Output of the System

CHAPTER 5 SYSTEM EVALUATION AND DISCUSSION

5.1 System Testing and Performance Metrics

The system will be tested on the hand detection module and the object detection module. Controlled environment is manipulated to simulate different condition and the performances of the system in different condition will be tracked. This is to help the system to be able to handle the scenario different conditions and flaws and failure will be examined to allow the issues to be fixed and improved in the future.

To evaluate the system performance, the system will be tested on different brightness condition, different types of background and number of objects and changing of targeting objects. Some of the weaknesses found and faced by the system will be discussed later in the project challenges and limitations. Besides that, the suggestions and improvements that can be made will also be proposed to reduce or fixed the issues that may arise as the system operates.

5.2 Testing Setup and Result

5.2.1 Testing on Different Brightness by Manipulating the Light Condition

The test has been done on the hand detection module in a controlled light condition which is divided into 3 categories, best brightness, moderate brightness and low brightness.

The testing is carried out in an enclosed room with 2 fluorescent light bulbs. For best brightness condition, all of the fluorescent lightbulbs are switched on and balanced light was applied. After that, one of the fluorescent lightbulbs will be switched off to create a moderate brightness condition and with an imbalanced light condition.

Lastly, to create a low brightness condition, all of the fluorescent lightbulbs will be switched off and only allow natural light from the sunlight. The test results were obtained and as shown below:

Case 1: Best Brightness with a balanced light condition

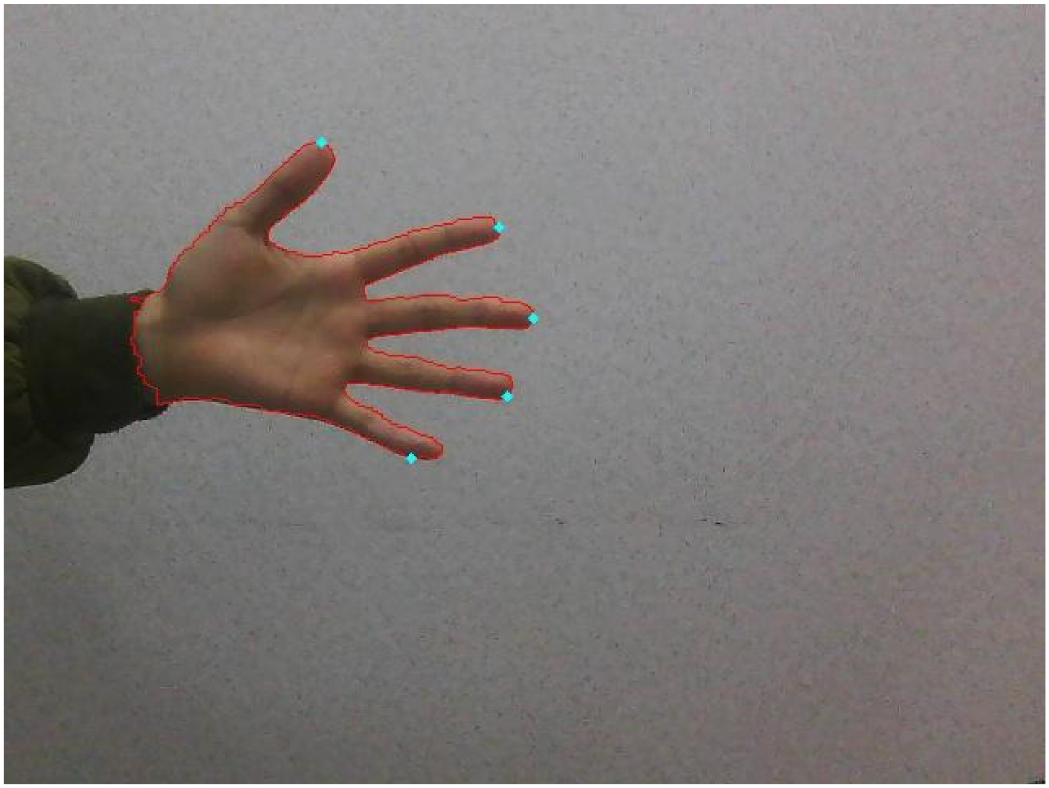


Figure 5.1 Test Result for Best Brightness Condition

From the result shown in Figure 5.1, the system is able to differentiate between the background and the user's hand. The location of the fingertip is precisely located on each of the fingers and marked with aqua blue points. The hand contour is clear and accurately divides the user's hand from the background.

The result can be obtained as above because the light condition is good enough to provide a piece of clean and sufficient colour information to the system and allow easier computation. Thus, the system can easily extract the user's hand out from the background and detects it.

Case 2: Moderate Brightness with Imbalanced light condition

Figure 5.2 Test Result for Moderate Brightness Condition

It is obvious to tell that the moderate brightness condition has affected the system on separating the user's hand with the background. As refers to the result above, the shadow is forming out on the background and the user's hand due to imbalanced light condition which causes the edge of the hand contour could not be precisely extracted. On locating and detecting the fingertip, the system could only detect 4 out of 5 of the fingers and this may be due to the noise that interrupt the system to detect the finger correctly.

From the result above, the system would easily misclassify the hand contour if shadow is present and background subtraction method could not easily differentiate between the shadows and the foreground object, thus, image segmentation is employed to further extract the region with skin colour in order to tackle moderate brightness condition.

Case 3: Low Brightness with only nature light condition

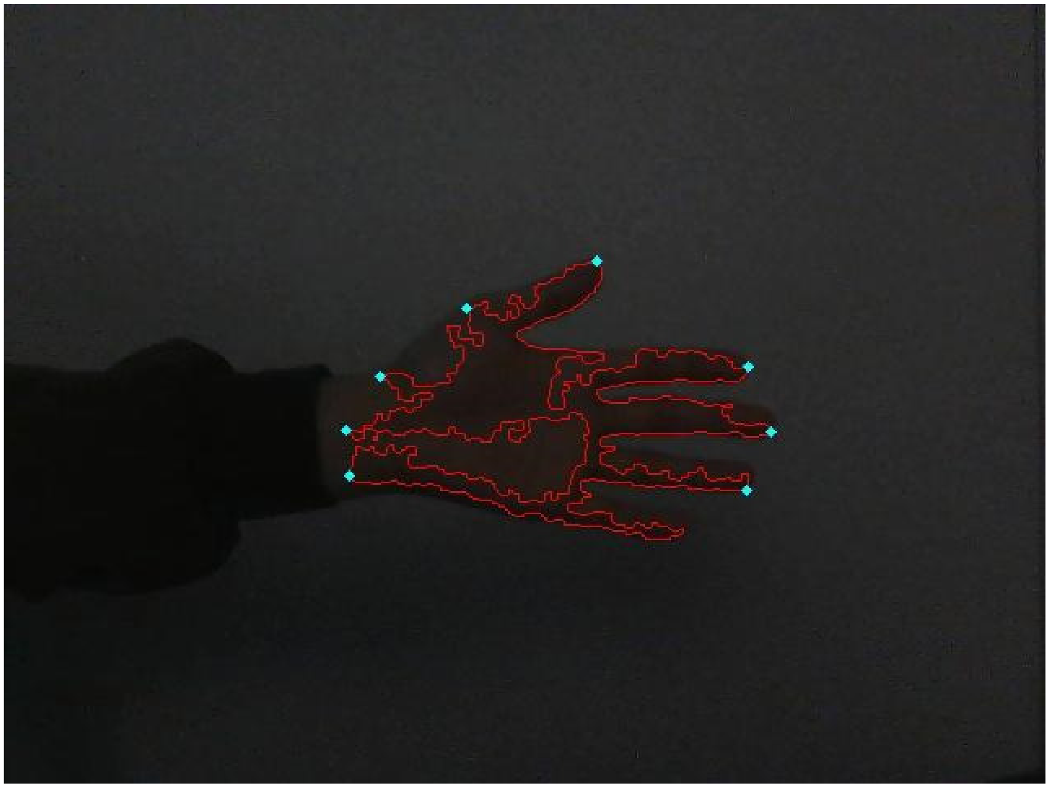


Figure 5.3 Test Result for Low Brightness Condition

According to the result obtain, the system has failed to precisely and correctly label out the edge of the hand contour and there is false positive detection on the fingertip detection. This is because the test was carried out in an enclosed room and limited sunlight are allowed to enter to provide natural light. The system fails to extract the colour information and the foreground information to differentiate between the user’s hand contour and the background.

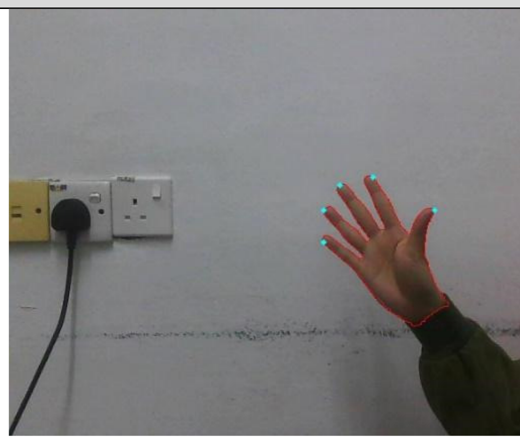
In low light condition, the system can be easily affected by the noise and shadow hence producing a false hand contour. The colour difference of the user’s hand with the background are also barely distinguishable. Therefore, it is reasonable to provide a proper light condition in order to make the system to work properly.

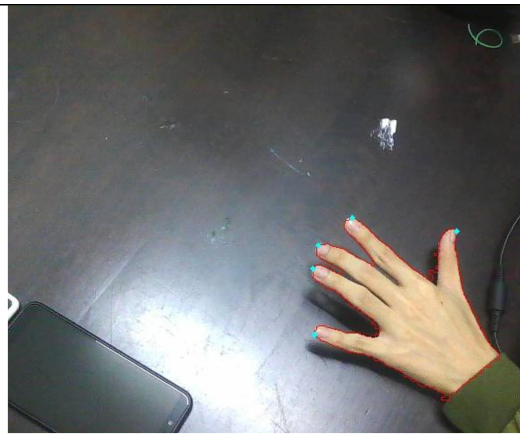
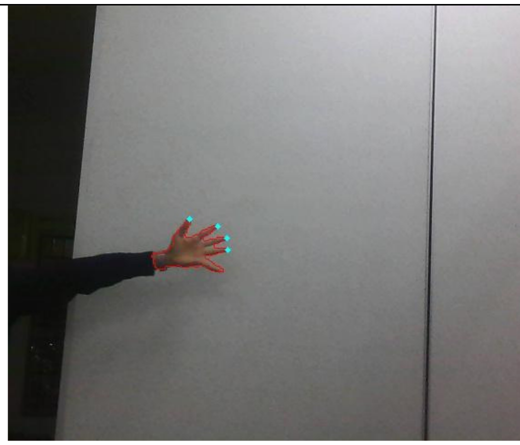
5.2.2 Testing on Different Types of Background

The system is built using background subtraction method, which means different types of background will indirectly influence the performance of the systems to locate the user's hand. The following testing is carried on 2 different type of background and let the system to differentiate the hand contour out from the backgrounds, which are the uniform background and the messy background.

To test the system on a uniform background, the background should have a plain and consistent colour with proper light condition. Some of the examples with a uniform background are the walls or a clean whiteboard. The user is required to move the hand to the front of a uniform background and the system will detects the user's hand. The test results obtained are shown as below. Noted that the detected fingertips will be labelled with aqua blue points in the test result.

Table 5.1 Test Results on Uniform Background

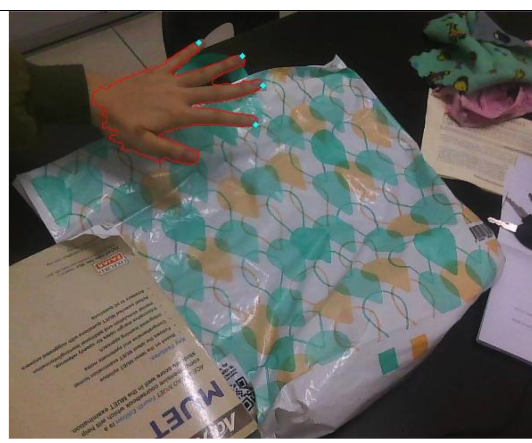
Test Case	Input Frames	Results
1		Number of fingers detected: 5/5

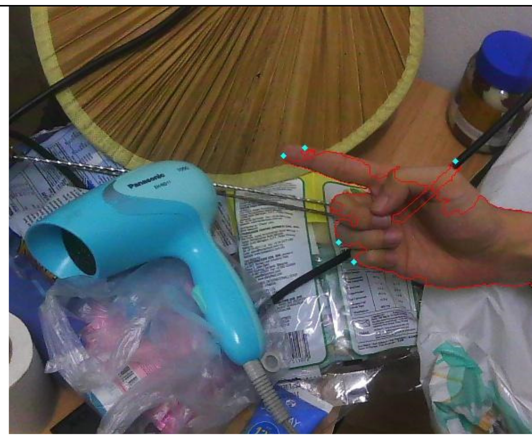
2		Number of fingers detected: 5/5
3		Number of fingers detected: 4/5 Reason: The hand is far away from the camera which reduce its accuracy
4		Number of fingers detected: 4/5 Reason: The fingertip edge is not clear enough

5		Number of fingers detected: 5/5
6		Number of fingers detected: 5/5
7		Number of fingers detected: 5/5

8		Number of fingers detected: 5/5

Table 5.2 Test Results on Messy Background

Test Case	Input Frames	Results
1		Number of fingers detected: 4/5 with false positive Reason: The hand edge is not clear
2		Number of fingers detected: 4/5 Reason: The hand edge is not clear

3		Number of fingers detected: 1/5 Reason: The hand is far away from the camera and the edge of the hand is unclear
4		Number of fingers detected: 1/5 with false positive Reason: Finger edges is not in a skin coloured background
5		Number of fingers detected: 5/5

6		Number of fingers detected: 4/5 Reason: The edge of the hand is not clear
7		Number of fingers detected: 5/5
8		Number of fingers detected: 4/5 with false positive Reason: Background colour is similar with the hand.

According to the test results, the system performs better in uniform background as compare to messy background. The reason behind is because the messy and clustered background contain more noise and edges that will interfere the systems to locate the hand correctly. Worst case scenario happens when the background has similar colour with the hands or the hand is far away from the camera.

5.2.3 Testing on Varying Number of Objects Class

The system the further tested on detecting varying number of objects presented in the background. This is to enable the performance check on the system ability to detects more common objects especially for the daily essentials. The greater the number of objects that can be detected, the better the system could provide meaningful information to the vision loss user. The test scenario is formed by putting some of the common daily essentials on a table and the system captures the input through the camera to detects the available objects that are on the table. The system needs to enclose the objects position on the output with a correct class name. The figure below shows the test results of the system output.

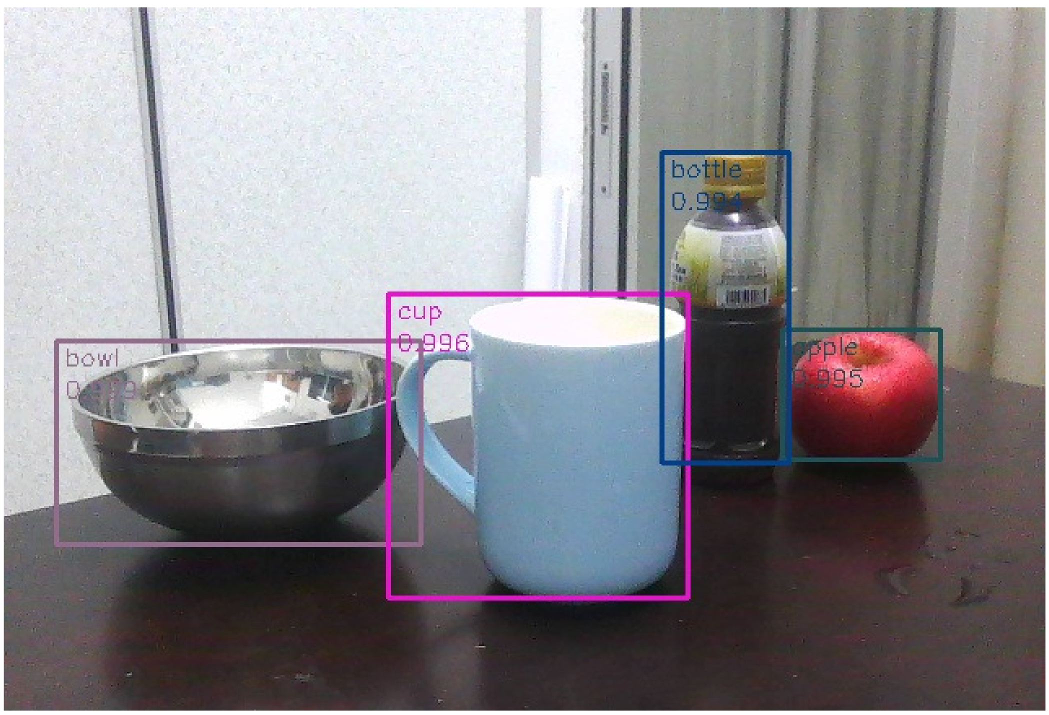


Figure 5.4 Test Result of Object Detection with Multiple Objects

The system successfully classifies the object in the correct position and their object class. This is because these objects have minimum overlapping region which allow they YOLO model to extract useful features to detects the object class. The testing is continued by adding more classes of object into the scenario and which create more overlapping and blocking between the objects. With the same settings as the previous testing (camera angle, light condition, etc), more detectable objects like the chair, fork and spoon are added into the scenario. The test result and the system output can be referred at the below.

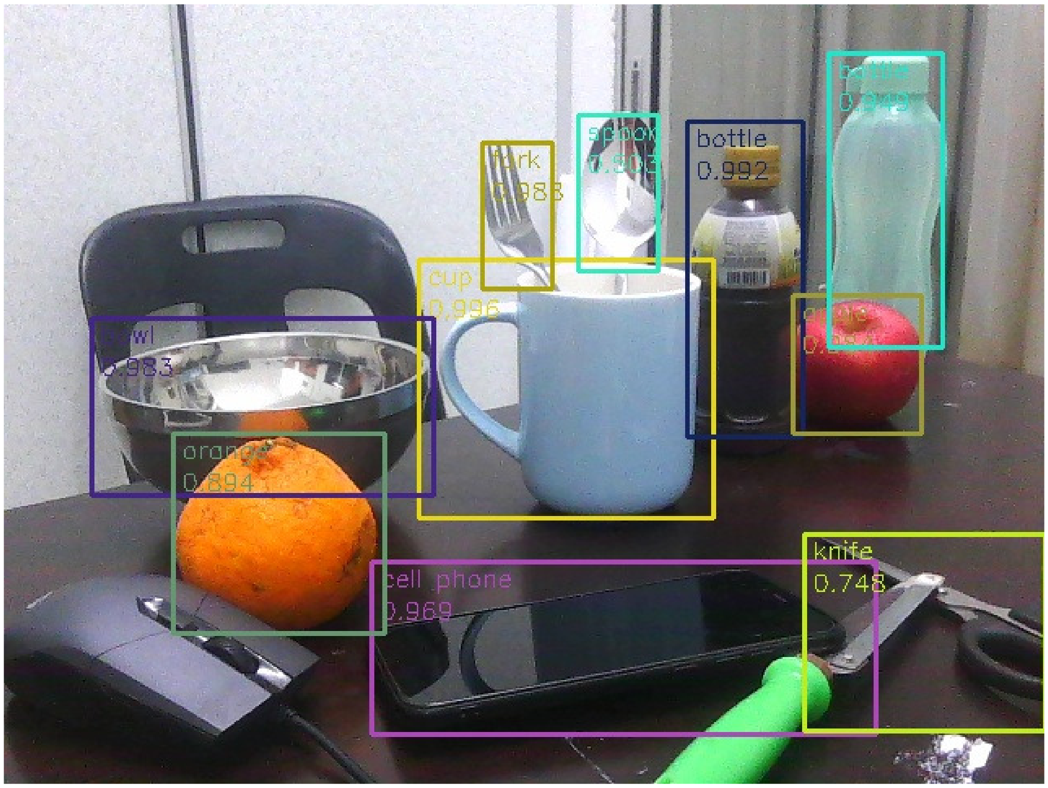


Figure 5.5 Test Result of Object Detection with Multiple Overlapping Objects

As can be seen from the result, the previously recognized objects can be successfully detected when more items are added to the scenario. The newly added items like the orange, cell phone, fork, spoon and another bottle are accurately detected and precisely labelled. However, more overlapping objects have led to some objects being misdetected and classified wrongly. The chair, mouse and the scissor are missed out by the object detection module and the knife is labelled in the wrong position. This is a mistake whereby the system might give wrong information to the blind user. Another similar test was performed to detect multiple overlapping objects that have the same object class. The test result is shown as the following.

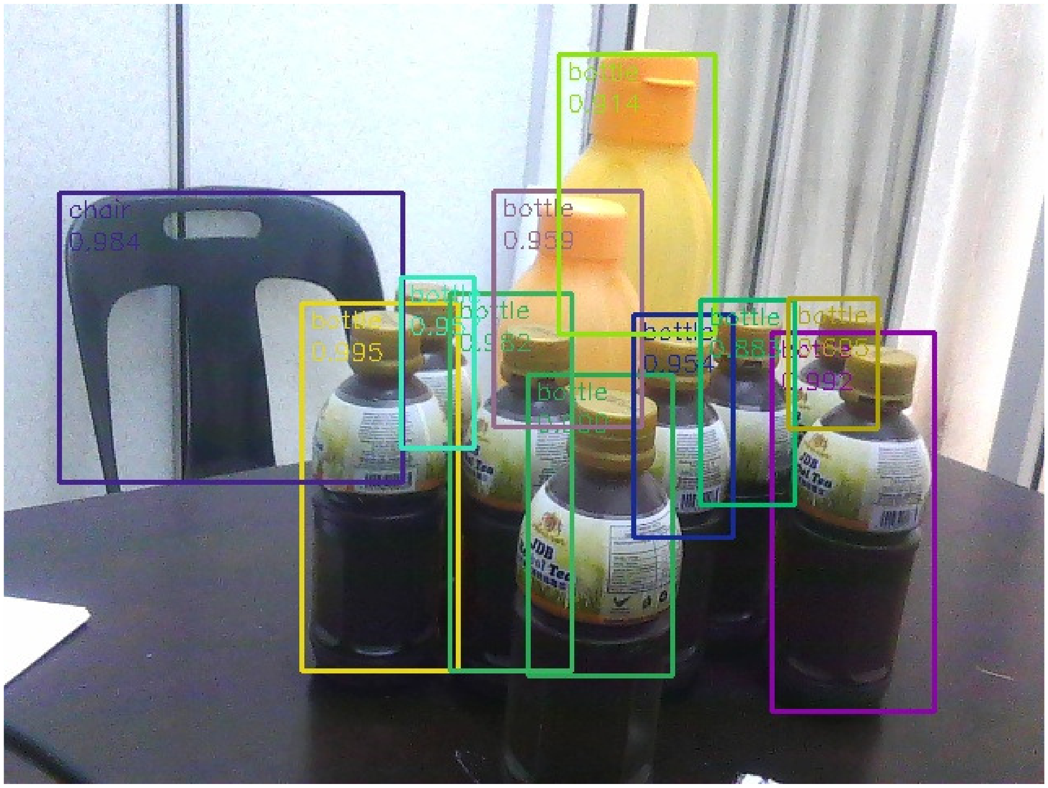


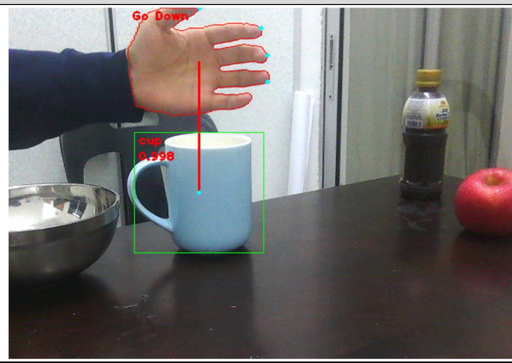
Figure 5.6 Test Result of Object Detection with Multiple Overlapping Objects with Same Object Class

The system was given with an input frame that contains multiple overlapping bottles (a total of 10 bottles) in the scenario. Although the system could detect the correct number of bottles from the input, there are a few faulty bounding boxes that label only the shown part of the bottle. The hidden and overlapped part of the bottle were missed out as a drawback of implementing the *NMSBoxes* function to remove redundant bounding boxes. Also, the blind user might get confused to get the desired "bottle" to target if there is multiple same class of object present in the scenario. The system should extract more interesting features such as the size or the colour of the bottle to distinguish between multiple same class objects to add to the system robustness.

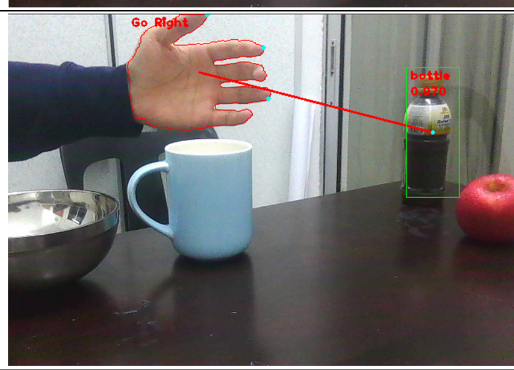
5.2.4 Testing on Changing Targeted Object

The system will detect the available objects from the camera feed and output the detection results. Multiple classes of objects are possible to exist in a different position in the scenario. The following testing is constructed to test the system's ability to differentiate between objects and give the correct instruction to the user to reach the object with their hand. The system should tell the direction of the hand of the user should move in order to reach a different kind of object that is present in the surrounding.

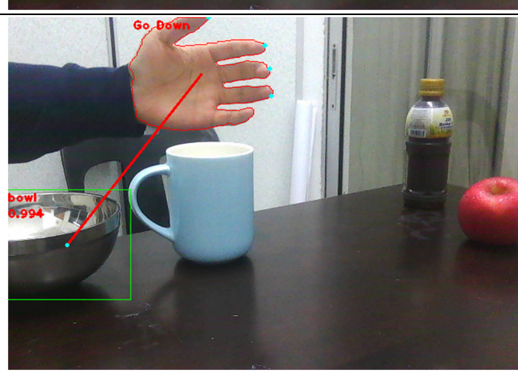
Table 5.3 Test Results on Different Targeted Objects

<table>
<tr><th>Test Case</th><th>Input Frames</th><th>Results</th></tr>
<tr><td>1</td><td>
</td><td>Targeted Object:
Cup

Direction to the Object:
Down

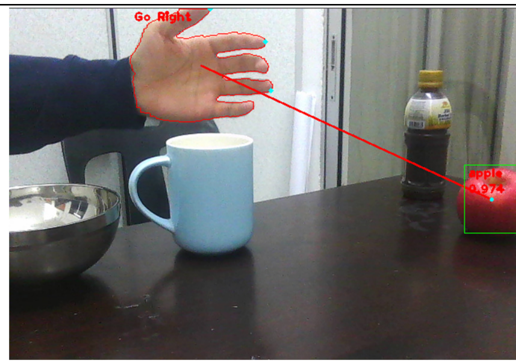
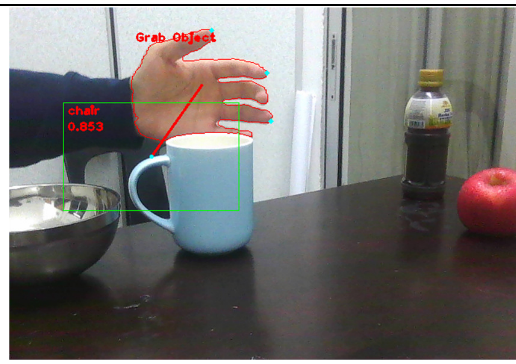
Instruction from the System:
Go down</td></tr>
<tr><td>2</td><td>
</td><td>Targeted Object:
Bottle

Direction to the Object:
Bottom right

Instruction from the System:
Go right</td></tr>
<tr><td>3</td><td>
</td><td>Targeted Object:
Bowl

Direction to the Object:
Bottom left

Instruction from the System:
Go down</td></tr>
</table>

4		Targeted Object: Apple Direction to the Object: Bottom right Instruction from the System: Go right
5		Targeted Object: Chair Direction to the Object: Backward in z-axis Instruction from the System: Grab object

From the testing above, the system could give a correct direction to reach a different targetted object. However, the system is unable to compute the spatial depth of the input frame from the camera and could possibly give the wrong instruction to the vision-impaired user. This is the weakness of the system as it fails to reconstruct the spatial information from the input image.

5.3 Project Challenges and System Limitations

The overall performance of the system is accurate and precise. Although the system could perform well in the best brightness and moderate brightness condition, the system will fail if the surrounding environment is lack of incoming light to provide a clear and bright input for the system to extract the information.

Along with it, the type of background will also affect the system performance because the system can easily get distracted by the messy and disarraying background as the system could fail to differentiate between the hand contour edge and the background edges. Having a background that has similar colour with the hand skin colour will further lead to misdetection of hand contour to happen.

Furthermore, the system will have difficulty detecting objects which are overlapping with one another because spatial information is limited in a 2D input frame. This also arises a problem in giving correct and accurate navigation instruction to lead the visually impaired people to locate the object in a 3D world. From the camera's point of view, the object may be located on the right but this is different for the people who are facing an opposite direction of the camera.

To overcome the issues above, the hand detection module could implement using artificial intelligence technology to train a detection model that could identify human hands. A better instruction can provide to the users if the system is able to recognize whether the moving hand is the left hand or the right hand. Also, the system could implement 3D scene reconstruction to further improve its usability and stability.

5.4 Objective Evaluations

As refers back to the project objectives, the system should be able to recognize objects, compute the direction and output the navigational guidance to the users in audio. By employing the YOLO object detection technique, the system is able to correctly identify common objects which are captured from the camera feed and label their position with the bounding box precisely. The system could still archive high detection rate on the overlapping objects if majority surface of the object is exposed.

Besides that, the system can successfully detect the user’s hand and compute the direction to reach a specific object. To ensure accurate instruction is given, the system can only target 1 object at each time to lead the user’s hand to grab the correct targeted object. This can be improved further by employing audio control module to inform the system to target another object without manually switching the targeted object.

Lastly, all of the instructions from the system will be output to the user in a window partition. To further improve the system usability, the text output is converted into auditory output with the help of Microsoft Speech API (SAPI Library) to output a computer-generated human voice in English to inform the user about the action to do in order to reach the objects.

CHAPTER 6 CONCLUSION AND RECOMMENDATION

6.1 Conclusion

People with visual impairment are having difficulty finding their objects and navigating the surrounding environment. They often require assistance and guidance from other people to understand the visual information on the surrounding environment. This has created a dependency of the visually impaired people on the other people and the assistive tools. Since they have an inability in receiving visual information, there is a complement to the problem which is by adopting computer vision technology to develop a system that aids in their vision in daily life. This paper has proposed an object finder system that can operate without the need of a human assistant, which supports the visually impaired to live independently. In addition, the proposed system is a cost-effective solution that can gradually reduce the financial burden faced by the vision loss people to afford expensive assistive tools. The development of the object finder system is necessary, as it would bring benefit to the visually impaired in seeking out a targeted object in an unfamiliar environment, especially by providing prior notice to them to avoid sharp objects like knives or scissors.

The system has successfully fulfilled the mentioned objectives, which are to develop an object detection module that detects indoor instruments, computes the direction to reach the detected object from the user's hand, and finally provides a navigational guide and instruction to the user to reach the targeted object with a computer-generated voice. To implement the requirements, the system has adopted a pre-trained object detection model which is the YOLO model. It can successfully identify objects and label with the object name and surround them with square boxes. The position of the boxes was used to compute the direction from the hand. Originally, the YOLO model does not detect the human hand hence the system archived it by using background subtraction and skin colour segmentation method to differentiate the human hand out from the background. Once the user's hand is detected, the system finds out the direction of the hand to reach the objects and generate auditory instruction for the visually impaired people.

6.2 Recommendation and Future Work

6.2.1 Improvement on Hand Detection

During the initial stage of the project development, the hand and finger detection module using the artificial intelligence approach has been tried but it did not work out. The current hand detection approach is accurate but it will somehow misclassify other types of motion as hand movement which lead to false positive. For improvement, the system could employ a hand tracking module from the MediaPipe library [14] which can accurately detect hand palms and perform finger tracking with a 3D hand landmark. This can help the system to detect and tracks finger movement in real-time with ease to improve the strength of the system.

6.2.2 Improvement on Object Detection

The current state of the project only runs the YOLO model on the background frame which is due to the fact that running YOLO object detection on every frame of the camera capture will heavily consume computational resources and cause latency and delay. The system can improve by adopting a temporal detection module which is to runs the YOLO model at a different time and frame intervals to minimize the computational consumption. This can help the system detects the object in real-time and allow more dynamic movement and change from the environment. In addition, the system can improve its functionality by allowing the user to register their personal stuff or friend to be detected. This requires retraining of the YOLO model which is time-consuming and can only be put into future development.

6.2.3 Improvement on Navigational Guidance

To allow the system to be more robust, Kinect technology could be implemented into the system to detect the depth of the environment. Various kinds of depth reconstruction and depth estimation could also be implemented to rebuild the environment depth and spatial information. With this, it allows the system to provide a more meaningful navigational guide to lead the users to the objects in a 3D environment. The system will be able to differentiate between left, right, up, down, and most importantly the front and the back to deliver a more accurate instruction.

REFERENCES

[1] D. M. Zuckerman, “Blind Adults in America: Their Lives and Challenges,” National Center for Policy Research for Women & Families, DC, United State. 2004. [Online]. Available: https://center4research.org/wp-content/uploads/2010/05/blind02041.pdf

[2] Centers for Disease Control and Prevention. “The Burden of Vision Loss,” Burden of Vision Loss | CDC. https://www.cdc.gov/visionhealth/risk/burden.htm (accessed Nov. 12, 2020).

[3] J. Brownlee, “A Gentle Introduction to Computer Vision,” A Gentle Introduction to Computer Vision. https://machinelearningmastery.com/what-is-computer-vision/ (accessed Nov. 5, 2020).

[4] E. Mohamed, “Deep Learning for Vision Systems,” Manning Publications. [Online]. Available: https://www.manning.com/books/deep-learning-for-vision-systems?query=deep%20learning%20for

[5] Z. Q. Zhao., et al., “Object Detection With Deep Learning: A Review,” in *IEEE Transactions on Neural Networks and Learning Systems*., 2019, vol. 30, no. 11, pp. 3212-3232.

[6] Csapó. Á., et al., “A survey of assistive technologies and applications for blind users on mobile platforms: a review and foundation for research,” in *J Multimodal User Interfaces 9,* 2015, pp. 275–286, doi: 10.1007/s12193-015-0182-7

[7] Redmon. J. and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Las Vegas, NV, USA, 2016, pp. 779-788.

[8] Hotaka. T. and Mayumi. A, “Assistive Systems for the Visually Impaired Based on Image Processing,” 2017, doi:10.5772/intechopen.70679

[9] Nishajith. A., et al., “SMART CAP – WEARABLE VISUAL GUIDANCE SYSTEM FOR BLIND,” in *International Conference on Inventive Research in Computing Applications (ICIRCA),* Coimbatore, India, 2018, pp. 275-278.

[10] Hoa. N., et al., “Web-based object detection and sound feedback system for visually impaired people,” in *International Conference on Multimedia Analysis and Pattern Recognition (MAPR),* Ha Noi, Vietnam, 2020.

[11] Kayukawa. S., et al., “BBeep: A Sonic Collision Avoidance System for Blind Travellers and Nearby Pedestrians,” in *CHI Conference on Human Factors in Computing Systems*, Glasgow, United Kingdom, 2019.

[12] Wang. H. C., et al., “Enabling Independent Navigation for Visually Impaired People through a Wearable Vision-Based Feedback System,” in *IEEE International Conference on Robotics and Automation (ICRA),* Singapore, 2017, pp. 6533-6540.

[13] Khamar. B. S, et al., “Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space,” in *3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015),* 2015, pp. 41-48.

[14] MediaPipe. “MediaPipe’s Hand,” Hands – mediapipe. https://google.github.io/mediapipe/solutions/hands.html (accessed Apr. 10, 2022).

APPENDIX

WEEKLY LOG

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Jan 2022	**Study week no.:** 1
Student Name & ID: Choo Yong Quan	
Supervisor: Prof. Leung Kar Hang	
Project Title: An Object Finder for the Visually Impaired	

1. WORK DONE

2. WORK TO BE DONE

- Discussion with supervisor on following plan
- Prepare FYP2 chapters and report
- Study on new FYP2 report structure and formatting
- Construct project plan

• Need to know how well YOLO version 3 can detect object types. Can send me a few pictures with results of the detection?

• Propose to have one fingernail painted red for easy detection in order to align your finger / hand with the object to pick up.

3. PROBLEMS ENCOUNTERED

4. SELF EVALUATION OF THE PROGRESS

Supervisor's signature

Student's signature

27 Jan 2022

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Jan 2022	**Study week no.:** 4
Student Name & ID: Choo Yong Quan	
Supervisor: Prof. Leung Kar Hang	
Project Title: An Object Finder for the Visually Impaired	

1. WORK DONE

- Detect object in red to align with the detection result
- Pictures with YOLOv3 detection results

2. WORK TO BE DONE

- Work on detect finger skin color
- Discussion with supervisor on following plan
- Prepare FYP2 chapters and report

Detect red color under different lighting conditions.

Build interface/demo to pick up the object when you close your eye under the condition that the distances from you to objects are within arm length

Build interface when the distances from you to objects are longer than arm length.

How to train YOLO to recognize new types of objects.

3. PROBLEMS ENCOUNTERED

4. SELF EVALUATION OF THE PROGRESS

Supervisor's signature

17 Feb 2022

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Jan 2022	**Study week no.:** 6
Student Name & ID: Choo Yong Quan	
Supervisor: Prof. Leung Kar Hang	
Project Title: An Object Finder for the Visually Impaired	

1. WORK DONE

- Detect object in red to align with the detection result
- Pictures with YOLOv3 detection results

2. WORK TO BE DONE

- Build interface/demo to pick up the object when you close your eye under the condition that the distances from you to objects are within arm length
- Build interface when the distances from you to objects are longer than arm length.
- How to train YOLO to recognize new types of objects.

3. PROBLEMS ENCOUNTERED

4. SELF EVALUATION OF THE PROGRESS

Supervisor's signature

Student's signature

3 Mar 2022

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Jan 2022	**Study week no.:** 7
Student Name & ID: Choo Yong Quan	
Supervisor: Prof. Leung Kar Hang	
Project Title: An Object Finder for the Visually Impaired	

1. WORK DONE - Detect object in red to align with the detection result - Pictures with YOLOv3 detection results
2. WORK TO BE DONE - Build interface/demo to pick up the object when you close your eye under the condition that the distances from you to objects are within arm length - Build interface when the distances from you to objects are longer than arm length. - How to train YOLO to recognize new types of objects.
3. PROBLEMS ENCOUNTERED - Finger detection contours not really well defined (currently using hsv detection to detect skin color)
4. SELF EVALUATION OF THE PROGRESS

Supervisor's signature

Student's signature

10 Mar 2022

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Jan 2022	**Study week no.:** 9
Student Name & ID: Choo Yong Quan	
Supervisor: Prof. Leung Kar Hang	
Project Title: An Object Finder for the Visually Impaired	

1. WORK DONE

- Detect object in red to align with the detection result
- Pictures with YOLOv3 detection results
- Tested out machine learning approach to detect human finger

2. WORK TO BE DONE

- Build interface/demo to pick up the object when you close your eye under the condition that the distances from you to objects are within arm length
- Build interface when the distances from you to objects are longer than arm length.
- How to train YOLO to recognize new types of objects.

3. PROBLEMS ENCOUNTERED

- Finger detection contours not really well defined (currently using hsv detection to detect skin color)
- Try to used machine learning method to detect fingertip but it does not work, is there any other detection method..

Should use motion detection to find finger.

4. SELF EVALUATION OF THE PROGRESS

Supervisor's signature

Student's signature

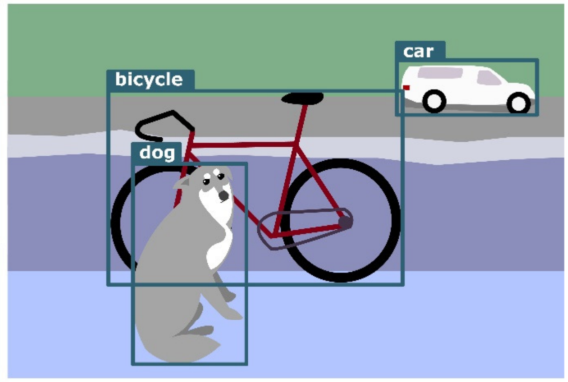
24 Mar 2022

POSTER

Universiti Tunku Abdul Rahman

Faculty of Information Communication and Technology

An Object Finder For The Visually Impaired



Introduction

The visually impaired people are facing difficulty in receiving visual information from the surrounding. By using advance computer vision technology, we can help the unfortunate people to regain the ability to 'see' the world again. This project proposes an object finder application that help the users to quickly identify and locate the surrounding objects as well as provide voice navigation guidance to reach the targeted object.

Objectives

The main objective of the project is to implement a real time object finder application that can work perfectly in the indoor environment. It can be divided into the following sub-objectives which include :-

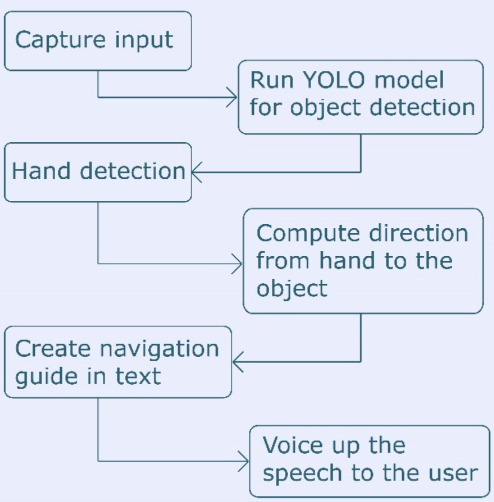
~ Detection of common indoor objects such as table, chair, cup and etc.

~ Detection of the user's hand

~ Compute the direction of the object from the hand to locate it

~ Voice command and navigational guide to reach the object

Methodology



Discussion

YOLO model are able to archive high speed and high accuracy in real time detection. We can use background substraction to detect the user's hand. Later, the system would compute the direction from the hand to the object. The system will issues a pick up notification when the hand overlapped with the object.

Work by: Choo Yong Quan

Supervised by: Prof. Dr. Leung Kar Hang

PLAGIARISM CHECK RESULT

ORIGINALITY REPORT

3%	1%	2%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Seita Kayukawa, Keita Higuchi, João Guerreiro, Shigeo Morishima, Yoichi Sato, Kris Kitani, Chieko Asakawa. "BBeep", Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019 Publication	<1 %
2	Hoa Nguyen, Minh Nguyen, Quan Nguyen, Shu Yang, Huy Le. "Web-based object detection and sound feedback system for visually impaired people", 2020 International Conference on Multimedia Analysis and Pattern Recognition (MAPR), 2020 Publication	<1 %
3	pdfcoffee.com Internet Source	<1 %
4	ijece.iaescore.com Internet Source	<1 %
5	Orita, Kazunori, Hotaka Takizawa, Mayumi Aoyagi, Nobuo Ezaki, and Mizuno Shinji. "Obstacle detection by the Kinect cane system for the visually impaired", Proceedings of the	<1 %

	Source	Similarity
	2013 IEEE/SICE International Symposium on System Integration, 2013. Publication	
6	Nadia Brancati, Giuseppe Caggianese, Maria Frucci, Luigi Gallo, Pietro Neroni. "Experiencing touchless interaction with augmented content on wearable head-mounted displays in cultural heritage applications", Personal and Ubiquitous Computing, 2016 Publication	<1%
7	www.ijarcs.info Internet Source	<1%
8	eprints.utar.edu.my Internet Source	<1%
9	www.scribd.com Internet Source	<1%
10	ai.googleblog.com Internet Source	<1%
11	laptopsunderbudget.com Internet Source	<1%
12	"Advances in Service and Industrial Robotics", Springer Science and Business Media LLC, 2020 Publication	<1%
13	Lecture Notes in Computer Science, 2015. Publication	<1%
14	machinelearningmastery.com Internet Source	<1%

Exclude quotes Off
Exclude bibliography On
Exclude matches Off

Form Title: Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)

Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	Choo Yong Quan
ID Number(s)	18ACB03979
Programme / Course	Bachelor of Information Systems (Honours) Information Systems Engineering
Title of Final Year Project	An Object Finder for the Visually Impaired

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceed the limits approved by UTAR)**
Overall similarity index: 3 % **Similarity by source** Internet Sources: 1 % Publications: 2 % Student Papers: 0 %	
Number of individual sources listed of more than 3% similarity: 0	
Parameters of originality required, and limits approved by UTAR are as Follows: **(i) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.***	

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

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

______________________________	______________________________
Signature of Supervisor	Signature of Co-Supervisor
Name: Leung Kar Hang	Name: ______________________
Date: 22 Apr 2022	Date: ______________________

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	1803979
Student Name	Choo Yong Quan
Supervisor Name	Prof. Leung Kar Hang

TICK (√)	DOCUMENT ITEMS Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
NA	Front Plastic Cover (for hardcopy)
√	Title Page
√	Signed Report Status Declaration Form
√	Signed FYP Thesis Submission Form
√	Signed form of the Declaration of Originality
√	Acknowledgement
√	Abstract
√	Table of Contents
√	List of Figures (if applicable)
√	List of Tables (if applicable)
√	List of Symbols (if applicable)
√	List of Abbreviations (if applicable)
√	Chapters / Content
√	Bibliography (or References)
√	All references in bibliography are cited in the thesis, especially in the chapter of literature review
√	Appendices (if applicable)
√	Weekly Log
√	Poster
√	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)
√	I agree 5 marks will be deducted due to incorrect format, declare wrongly the ticked of these items, and/or any dispute happening for these items in this report.

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

I, the author, have checked and confirmed all the items listed in the table are included in my report.

____
(Signature of Student)
Date: 22 April 2022