Enhanced Camera Target Tracking Using a Low-cost Thermal Camera

Presented by:
Lezerick Graeme Owies

Prepared for:
J. Son
Dept. of Electrical and Electronics Engineering
University of Cape Town

Submitted to the Department of Electrical Engineering at the University of Cape Town in partial fulfilment of the academic requirements for a Bachelor of Science degree in Mechatronics (Electrical Engineering; Mechatronics or Electrical and Computer Engineering)

July 21, 2023

Declaration

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2. I have used the IEEE convention for citation and referencing. Each contribution to, and quotation in, this report from the work(s) of other people has been attributed, and has been cited and referenced.
3. This report is my own work.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as their own work or part thereof.

Signature:.........................
L. G. Owies

Date:.............................

Acknowledgments

I would like to acknowledge the following people for helping me complete this thesis in a timely matter by providing me with support and advice over the duration of this project.

- I would like to thank my supervisor Jarryd Son who set out time to meet with me on a weekly bases to ensure that every thing is on track. I would also like to thank him for taking time to respond to my MS Teams messages and for providing valuable feedback on how to improve my literature review. Lastly I would like to thank him for providing me with resources like 3d printing and a Raspberry Pi without charge.
- Special thanks to Sergio Canu from Pysource and Joshua Hrisko from MakerPortal as their tutorials and source code really helped me get a firm understanding of how to use OpenCV for object tracking as well as setup the thermal camera.
- I would also like to thank Brendan Daniels at white lab for helping me 3d Print my mechanical parts.
- Lastly I would like to thank my family for supporting me not only over the duration of this thesis but over the 4 years of my degree. A special thanks goes to Dr. Jacob Cloete who has given me advice on how approach certain scenarios and guided me when I needed it.

Abstract

Visual cameras used in object tracking applications often struggle in poor visual conditions sch as dark environments. This report investigates integration of a low-cost thermal camera into the design of visual tracking systems to improve their performance under poor visual conditions. This high level objective was then broken up into smaller sub-objectives which aided with design processes such as component selection, mechanical part designs and software development. All components were selected to provide the best results within budget constraints with main components being the Adafruit AMG8833 Thermal camera, Raspberry Pi 5MP Mini camera and the Raspberry Pi 3B. The mechanical parts were designed after the component selection process and the dimensions were tailored to house the specific components. Two algorithms were developed, one that uses only the visual camera to perform tracking and another one that uses both cameras. The first algorithm was used as a benchmark to see how much the use of a thermal camera can improve the performance of the system. The use of a thermal camera allowed the device to track objects in darker environments where the visual camera performed poorly. However the overall real-time tracking performance of both algorithms was poor due to the computational limits of the Raspberry Pi 3B. Overall the device met all the basic requirements set out at the start but can be improved a lot in future work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background to the study

Camera tracking systems have become more popular in recent times with the rise of computer vision in the main stream along with the availability of better computational hardware at cheaper prices. The most popular modern applications include self-driving cars and home security and surveillance systems.

Typically these systems would make use of a visual camera and a some sort of computer vision object detection algorithm to identify and track an object on a visual frame. The most common algorithms used for these tasks are CNN based algorithms such as YOLO, RCNN and Fast-RCNN. Other object detection methods like background subtraction are also used to track moving targets when the project has hardware limitations and the background of the visual frame is constant over time.

A common problem with using visual cameras for object detection is that they struggle in poor visual conditions especially in dark environments. Thus the aim of this thesis is to investigate how the integration of a low cost thermal camera into visual tracking systems can improve the overall performance in poor visual conditions.

1.2 Problems Statement

The use of only visual cameras in tracking applications (i.e, self driving cars and security surveillance systems) are limited to situations with good visual conditions. As a result such systems are rarely able to be used in the night or in enclosed areas where there is little to no light.

1.3 Project Objectives

This project aims to improve the performance of visual camera tracking systems under poor visual conditions by incorporating a low-cost thermal camera into the design using sensor fusion techniques. To achieve the main objective the project can be subsequently divided into smaller sub-objectives. These sub-objectives include:

- The design and testing of a camera tracking system using only a visual camera
- The design and testing of a algorithm that improves the performance of the visual tracking system under poor performance by incorporating a thermal camera
- The design and construction of a mechanical platform that can rotate the cameras with at least one degree of freedom in order to keep a target within the field view

1.4 Terms of Reference

The table below outlines the terms of reference for this project.

Table 1.1: Software subsystem acceptance test protocols

Project ID:	JS-03
Supervisor:	Jarryd Son
Description:	Visual cameras are commonly used sensors for a variety of applications such as pedestrian tracking in self-driving cars and surveillance. Visual cameras suffer under low visibility conditions which could be caused by poor lighting, fog etc. This project aims to develop a camera tracking system that can be used to keep a target within the visual field. Additionally, the project aims to use a low-cost thermal camera as an additional source of data that should improve the performance of the visual target tracking system. The thermal images would be able to provide some useful information regardless of lighting and environmental conditions and should improve the performance of the overall target tracking system.
Deliverables:	The student is expected to complete the following: • Refine the technical requirements of the proposed project and develop acceptance test procedures • Design and implement: – A visual tracking system to track at least a single target (e.g.person) in a visual scene – An algorithm to incorporate the thermal image data to improve the visual tracking under poor visual conditions – A mechanical platform to move the cameras to track the target and keep it within the visual field. A single degree of freedom will suffice for this platform • Perform test procedures on sub-systems and integrated solution. • Submit a final report
Skills/Requirements:	This project requires skills in electronics and embedded systems. Machine learning knowledge may be beneficial.

GA1: Problem Solving: Identify, formulate, analyse and solve complex engineering problems creatively and innovatively	This project aims to develop a target tracking camera that incorporates visual and thermal images. The combination or potential fusion of information from these two sources is non-trivial and requires innovative problem-solving attributes. This project requires identifying or formulating appropriate algorithms for effectively utilising both sources of data (visual and thermal images), selecting hardware, and developing a moving platform as necessary
GA 4: Investigations, Experiments and Analysis: Demonstrate competence to design and conduct investigations and experiments.	Design: Refinement of technical requirements, development of acceptance test procedures, justifiable selection of hardware for the task, design of the algorithms to incorporate both data sources Investigations: Investigate visual target tracking and the use of additional supplementary data sources, especially thermal images. Experiments: Analyse and compare the performance of the visual tracking with and without incorporation of the thermal image data
Extra Information:	Link: *Infrared and visible image fusion methods and applications: A survey* • Most of these are beyond the scope of the project, but this should at least provide some context
BROAD Research Area:	Visual tracking
Ethics:	No
Project suitable for ME/ECE/EE/ALL:	ME

1.5 Scope and Limitations

1.5.1 Scope

The scope of this project includes the design and testing of visual camera tracking system along with the integration of a thermal camera for improved tracking under poor lighting conditions. A detailed scope breakdown can be seen below.

- Design and implementation of an object tracking system using a visual camera
- Development of sensor fusion algorithm for thermal- and visual-camera
- CAD design and construction of a mechanical platform to mount components
- The development and implementation of suitable Acceptable test protocols for each subsystem along with the final system

The scope of this project does not include the following:

- Fast real time performance of the camera tracking system

1.5.2 Limitations

This project had the following limitations:

- A limited budget of R2000
- Time constraint of 14 weeks
- Limited to a Raspberry pi 3B which has insufficient hardware to execute the project properly (Lack of a GPU)
- Limited to real-time test data i.e, no simulations
- Limited to a quality thermal camera

1.6 Plan of development

The report remainder of the report has been structured as follows:

Chapter 2 - Literature review - Provides insight into literature for the concepts used during this project

Chapter 3 - Methodology - Describes methodology followed in this project

Chapter 4 - Design - Provides a detailed design process of the system

Chapter 5 - Results - Provides the results obtained from testing the system

Chapter 6 - Conclusion and Discussion - Discusses the results obtained, provides concluding remarks and future recommendations

Bibliography

Appendices

Chapter 2

Literature Review

This section aims to provide insight into relevant literature associated with the implementation of the given task. Its starts out by taking a closer look at thermal cameras which are one of the primary sensors needed for this project. It then takes a look at sensor fusion which is one of the primary objective of this project. Finally it takes a look at object detection algorithms which are used to detect humans on a given frame to allow for tracking to be implemented.

2.1 Visual cameras and object detection

A visual camera can best be described as a device with the capabilities to capture and store visual information. The earliest form of a camera as we know it in the modern era was created by Joseph Nicéphore Niépce in early 1800s. He managed to capture the first permanent photograph that didn't fade easily [16].

Most cameras used today are digital and are present in many commercial devices such as self-phones, surveillance systems and laptops. Digital cameras are comprised of a 2d-array of millions of light sensitive photosites that store information in the form of an electrical signal when exposed to light. The strength of the electrical signal generated by each photosite is then measured and processed to determine the color of the pixel and subsequently form the digital image [17]. The amount of photosites in the camera lens define what is commonly known as the megapixel rating of the camera with each pixel being comprised out of three photosites [17].

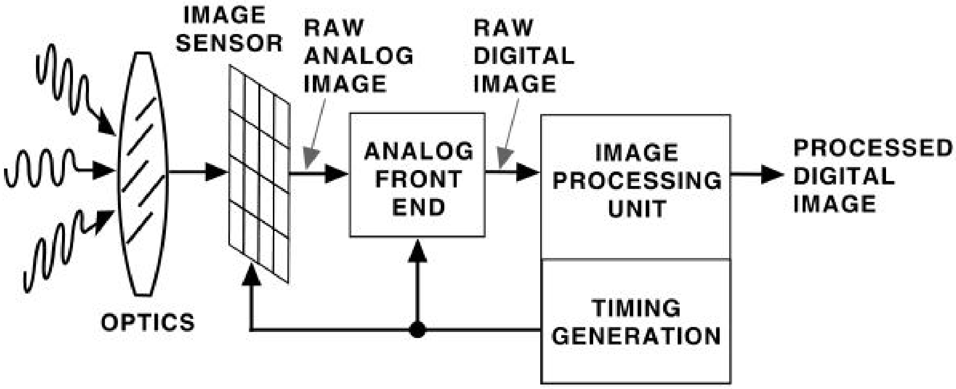


Figure 2.1: Basic diagram showing how a digital camera generates a digital images from incoming light [1]

Even though digital cameras are excellent for capturing still images and videos in most scenarios they still have significant performance issues when used for object detection applications. The most common problem is that visual cameras need to have an external light source present in order to detect objects in the environment. Even a slight change in the illumination level of the environment can severely impact the performance an object detection algorithm due to objects appearing different under varying light conditions. Another problem with visual cameras is occlusion which means that objects can sometimes be partially hidden behind other objects that are in the foreground of the frame and as a result they become difficult to detect [18].

To improve the performance of object detection under low light conditions Morawsk et al. [2] proposed the used of image enhancement methods and patch -wise light augmentation along with object detection algorithms. However they were only able to achieve a slight improvements in detection compared to using the object detection models without their proposes improvements as seen in figure. The authors also stated that there currently exists a huge gap in performance of object detection algorithm when they are used in extreme low light conditions compared to when they are used in non-extreme conditions [2].

Detector	init.	Proposed Enh. Module	AP_{50}	AP_{75}	AP
RetinaNet [[illegible]]	COCO [[illegible]]	✗	72.8%	47.6%	45.9%
		✓	**74.1%**	**49.1%**	**47.5%**
PAA [[illegible]]	COCO [[illegible]]	✗	71.6%	47.5%	45.4%
		✓	**73.0%**	**48.5%**	**46.7%**
Faster R-CNN [[illegible]]	ImageNet [[illegible]]	✗	61.6%	34.5%	33.3%
		✓	**64.2%**	**35.8%**	**35.5%**
FCOS [[illegible]]	random	✗	56.2%	**22.1%**	26.4%
		✓	**58.1%**	21.0%	**26.7%**

Figure 2.2: Performance comparison of popular object detection models under extreme light conditions with and without image enhancements light [2]

2.2 Object Detection and Tracking

Object detection is a process within the realm of computer vision that aims to identify and locate specific objects in a digital video or image. Object detection is often the first task performed in many computer vision processes as it provides crucial information about the detected object. The information gathered from the object detection stage allows computer vision systems to perform several tasks including feature recognition and object tracking and as a result it is used in many modern applications such as surveillance systems and security systems, robotics and autonomous driving vehicles [19].

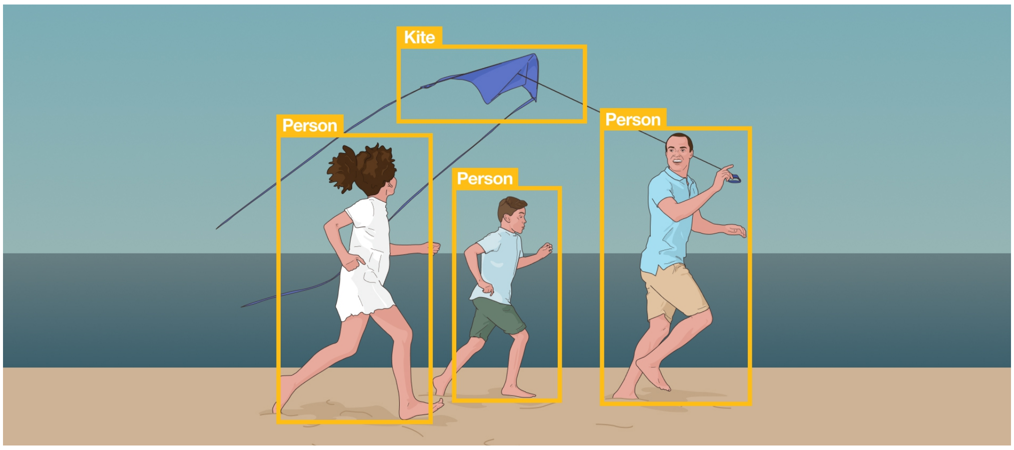


Figure 2.3: illustration of how an object detection algorithm identifies different objects on a visual frame [3]

2.2.1 Convolutional Neural Networks (CNN)

Convolution Neural Networks are artificial neural networks that are typically used in image processing and recognition tasks. Most modern object detection algorithms such as Region based Convolutional Neural Networks (R-CNN), Fast R-CNN and You only look once(YOLO) make use of the CNN architecture as a base for performing detection [20]. The CNN Architecture normally consist of three types of layers including the convolution layer, pooling layer and fully connected layers. Each of these layers perform a specialised task that is critical in the success of the algorithm [14]. A detailed breakdown of each layer of the CNN can be found in Appendix A.1

2.2.2 Object Detection Algorithms

Object detection algorithms can be classified into two classes namely single state-stage detectors and two stage detectors. The most common single-stage detectors include the Single Shot Multi-box detector(SSD) and YOLO algorithms while the most common two-

stage detectors include R-CNN, Fast R-CNN and faster R-CNN [21].
Two stage detectors work by first identifying a region of interest (ROI) on the given image and then feeding the ROI into a secondary stage that does object classification and boundary box regression. Single-stage detectors take a much simpler approach by using simple regression methods to classify and locate the given objects on a image. The main difference between single-stage and two-stage detectors is the trade of between speed and accuracy. Single stage detectors are generally faster than two-stage detectors which generally have better accuracy [21].

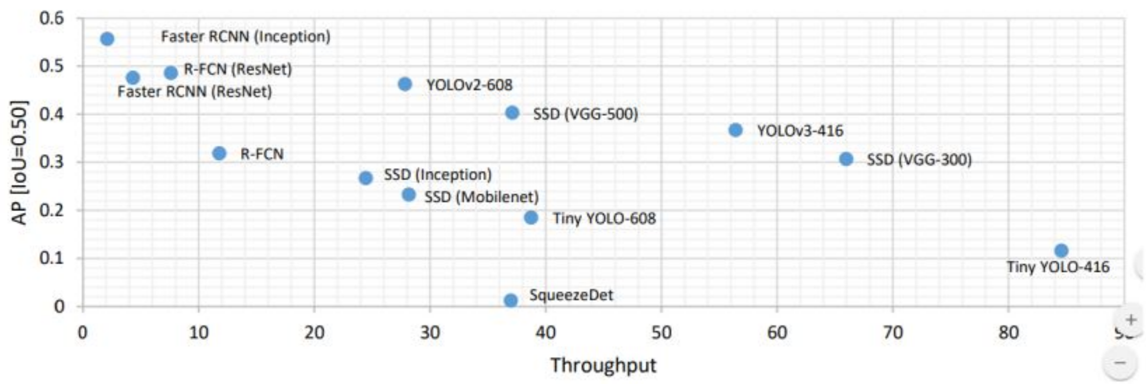


Figure 2.4: Average inference time vs Average Accuracy of various Object detection models (Tested on NVIDIA Jetson Nano) [4]

Due to the superior speed in single-stage detectors they are often considered for applications that have limited hardware resources i.e, embedded systems. However it is worth noting that the increase in speed may severely impact the accuracy to the point where the obtained results are sub par. Thus, Kim et al.[4] proposed a study to find the object detection algorithm that possesses the optimal accuracy and speed trade-off for deployment on embedded systems. Their study found that YOLO-V3 performed the best within the given criteria with Faster R-CNN and YOLO-v4-tiny producing the best accuracy and speed respectively. The results of their study can also be seen on figure 2.4 above.

A more traditional method of object detection is background subtraction. Background subtraction is a simple object detection technique that works by comparing the input image to a similar reference frame that doesn't contain any objects. The reference image pixel values are subtracted from the input image pixel values and then fed into a pre-defined threshold to determine if there is an object on the frame [5].

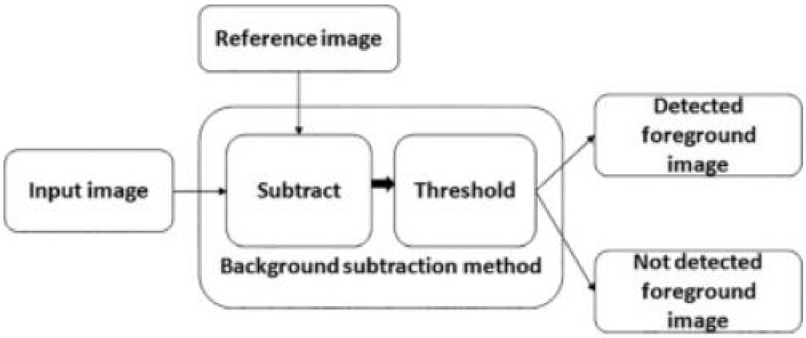


Figure 2.5: Basic structure of background subtraction algorithm [5]

Background subtraction techniques can be mainly classified into two classes namely recursive and non-recursive techniques. Non recursive techniques work by storing previous input frames and using them estimate to objects on future frames. The most common non recursive techniques include frame differencing, frame differencing, median filtering and linear predictive filter. Recursive techniques work by recursively updating a single reference background model based on current input frames. Recursive techniques include the approximated median filter, kalman filter and mixture of Gaussian (MoG) methods [22]. In general recursive techniques require less hardware resources compared to non-recursive techniques as non-recursive techniques needs to store information which consumes resources [23].

2.3 Thermal Imaging and Infrared Cameras

2.3.1 Brief History

The existence of infrared light was first discovered by German astronomer Sir William Herschel in the 1800s. He made his discovery whilst performing an experiment where he passed sunlight through a glass prism to split it into various colours with the aim of determining how much heat is contained within each colour on the visible light spectrum. During his experiment he noticed that the amount of heat contained in each colour increased as he moved from violet to red, thus out of curiosity he decided to measure just beyond the visible red and found that this area was even hotter which lead to the discovery of infrared light [24]. Sir Williams discovery ultimately led to the first form of a thermal camera being developed by Hungarian physicist Kalman Tihany for the British military in 1929. In the latter half of the 20th century American companies Texas Instruments, Hughes Aircraft and Honeywell made further improvements on Tihany's invention and essentially laid the foundation for the wide spread application of thermal cameras and thermal imagery today [25].

2.3.2 Brief Theory

Thermal cameras make use of infrared radiation emitted from objects to create an image. Infrared radiation occupies the area just below visible red light on the electromagnetic spectrum and has a wavelength ranging from approximately 780nm to 1mm [26].

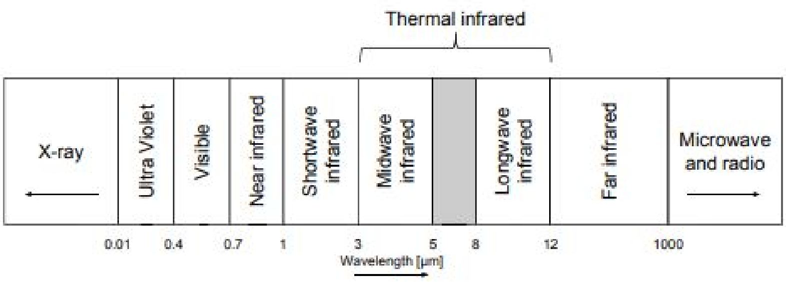


Figure 2.6: A diagram indicating the position of infrared radiation on the electromagnetic spectrum [6]

The infrared section of the electromagnetic spectrum can be further sub-devised into various categories including Near Infrared (NIR), Short wavelength Infrared (SWIR), Medium Wavelength Infrared (MWIR),long wavelength infrared (LWIR) and Far Infrared (FIR). However the FIR,SWIR and NIR sections do not form part of the thermal infrared spectrum which is the range of wavelengths used for thermal imaging, thus the the thermal infrared spectrum occupies the range of wavelengths between $3\mu m$ to $12\mu m$[6]. Physically thermal cameras make use of an array of sensors called a focal plane array to detect infrared radiation, the sensors then produce an electrical signal is relayed to computer/micro-controller to produce an image. The quality of the image produced is directly related to the amount of sensors used as well as their sensitivity. As different temperatures produce different thermal radiation the hardware used in any given application can be tailored to be more sensitive within a given range of infrared light for a known situation [7][27].

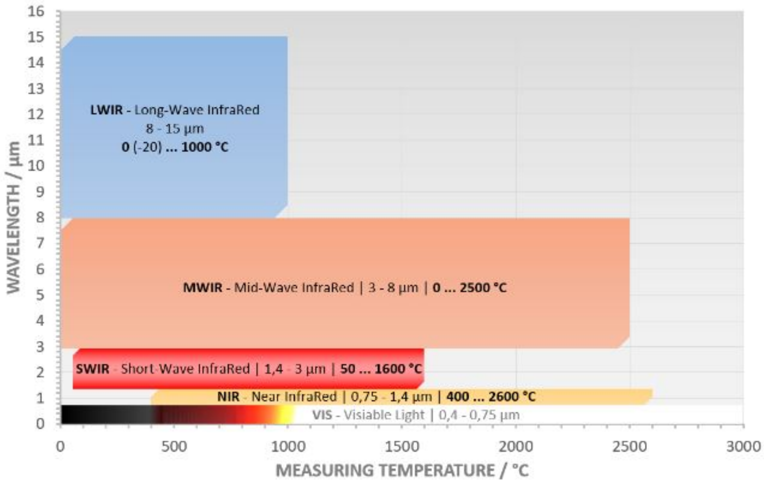


Figure 2.7: A Diagram diagram indicating which range of infrared radiation works best produce the best results for detecting specific temperature ranges) [7]

A detailed breakdown of the physics of thermal cameras can be found on Appendix A.2.

2.3.3 Modern Application

Thermal cameras are used in a variety of modern day applications due to the unique perception they provide of the environment. Common uses include the maintenance of electrical wiring and equipment, Mechanical installations, security and surveillance systems, Gas defections and a variety of medical procedures which help with the detection of breast cancer,diabetes and nephropathy and vascular disorders [28] [29]. With the rise of AI and computer vision over the past decade accompanied with the commercialisation of thermal cameras it has also seen them being applied within the realm of object detection and tracking [30].

2.4 Sensor Fusion

Sensor fusion is the process of merging information from two or more sensors to create a more accurate representation of reality and subsequently improve the quality and robustness of the overall system. Due to the increase in sensors available on the market at various prices and varying functionality the use of sensor fusion has become more popular in recent times.

According to Prajapati et al.[31] the use of sensor fusion provides four general performance improvements. These performance improvements can be seen through a higher abstract level of the data, improved certainty and reliability of the system, improved accuracy and noise reduction of the data as well as improved completeness of the data. However a key issue with sensor fusion is that individual sensors often tend to have different data structures and formats in which they provide data. The conversion of the source data to a common reference frame is crucial to the success of the sensor fusion process [31][32].

In general sensor data can be combined in three fundamental ways namely competitive fusion, complementary fusion and co-operative fusion. Each of these methods aim to improve a specific aspect of the system and thus in most applications a combination of all three methods is used [8].

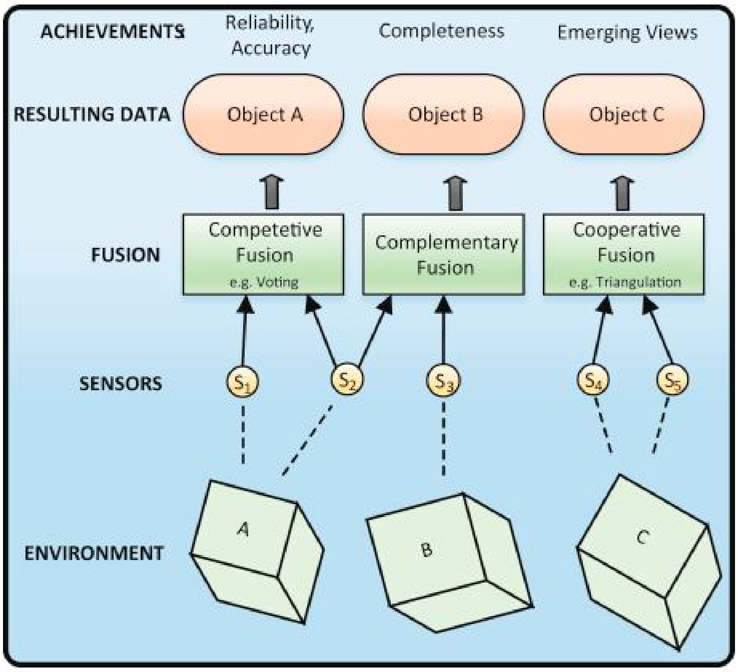


Figure 2.8: A Diagram diagram indicating the three fundamental ways of data combination for sensor fusion and the improvements they make to they make to the system [8]

Thermal imagery has brought a new dimension to object detection and tracking as they are less sensitive to environmental changes compared to visual cameras, thus making thermal cameras better suited for certain applications. However due to the limitations of thermal cameras the idea of using sensor fusion to combine visual- and thermal-cameras to increase the performance of object detection systems is one that is often proposed in the literature.

In a study done at Linköping University researchers found that the use of sensor fusion between visual- and thermal-cameras for object detection under varying lighting and weather conditions outperformed the cases where the sensors were used individually [23]. Sensor fusion can however also be detrimental to the performance of the system if it isn't applied properly. This was demonstrated in a study done by Abidi et al.[9] in which they attempted to use visual and thermal cameras to improve facial recognition software by making use of different sensor fusion techniques. During the study they found that only the decision based sensor fusion method consistently produced better accuracy in all scenarios than the use of the individual sensors. The results from their can also be seen below in figure 2.9

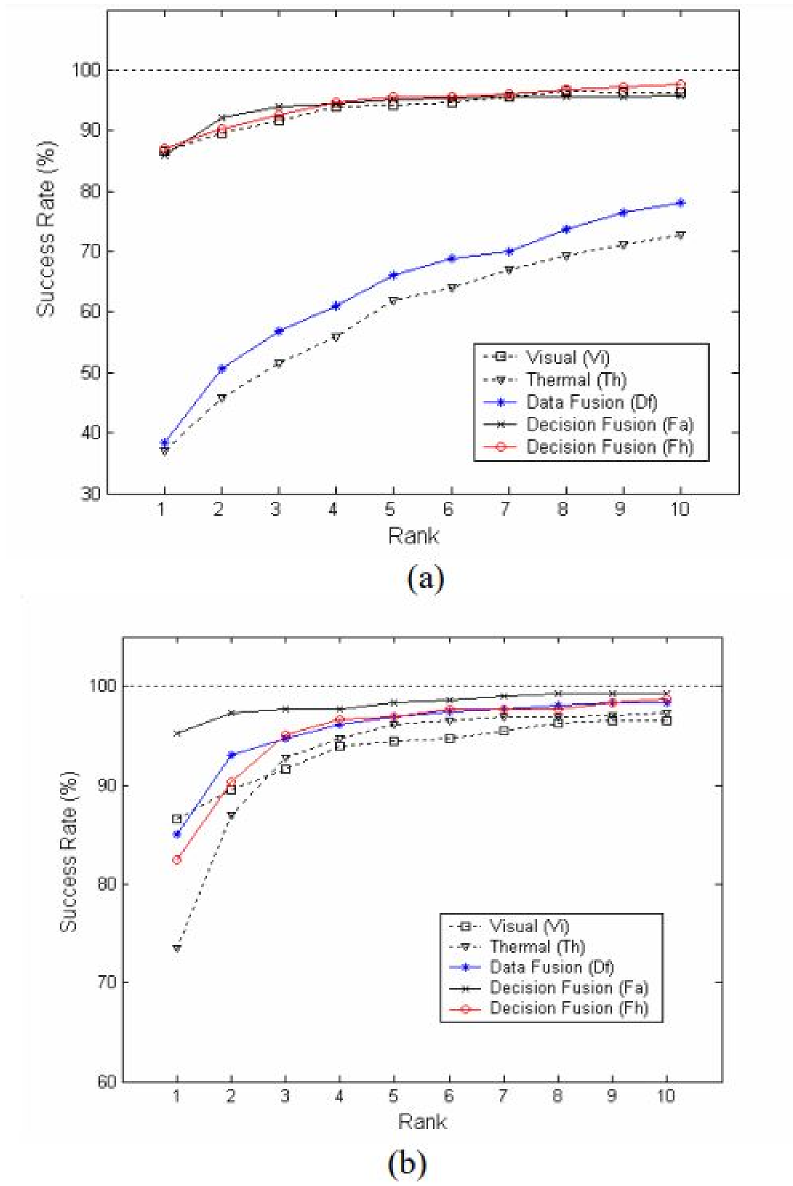


Figure 2.9: Comparison of different sensor fusion methods and their accuracy in two different scenarios [9]

2.5 Concluding Remarks

Through investigating the literature discussed in this chapter I have found that camera tracking systems that make use of only visual cameras can be severely impacted by poor lighting conditions and object occlusions due to the limitations of digital cameras. Morask et al. [2] proposed methods to improve the performance of object detection algorithms under extreme low light conditions but were only able to achieve a very slight improvement. This reaffirms the notion that the problem lies with visual cameras and their limitations rather than object detection models.

Current object detection models are mostly CNN Based a with the most popular ones being Yolo, SSD, R-CNN, Fast R-CNN and Faster R-CNN. Out of these Faster R-CNN is the most accurate and Yolov4-tiny is the fastest. According to Kim et al. [4] the best model to use on devices with limited computational hardware is YoloV3 as it has the most optimal speed-accuracy trade off. However I think in situations where real time performance is key and the device has limited computational resources YoloV4-tiny is the best due to the high inference time. It would also be beneficial to further investigate the use of object detection algorithms on embedded devices in order to develop more efficient models.

The use of sensor fusion between visual and thermal cameras has also been proposed in the literature by Abidi et al and Bergenroth. In both studies the use of sensor fusion between the visual and thermal camera improved the overall systems performance, However Abiddi et al [9] showed that sensor fusion can also be detrimental to the performance of the system in implemented incorrectly. This thesis also aims to use sensor fusion techniques to improve the performance of object detection under poor visual conditions, however it will make use of a low resolution thermal camera which brings along new challenges to overcome.

Chapter 3

Methodology

3.1 High Level Description

As mentioned in Chapter 1 the aim of this project is to produce a camera tracking system that incorporates a visual- and low cost thermal camera. To help achieve this goal the project followed a V-Diagram design methodology as it is a simple model that allows for sufficient testing of subsystems and the final design.

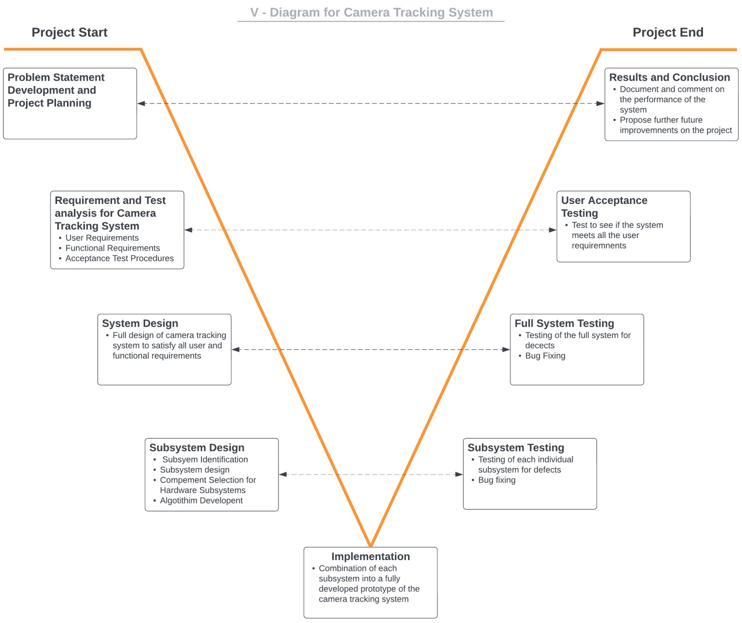


Figure 3.1: V-Diagram showing how the planned project progression

Based on the problem statement in Chapter 1 the user requirements were extracted and refined into functional requirements. This is an important step in the design process as it ensures that the final design meets all the users demands, whilst also providing a way of breaking the system into various hardware and software subsystems.

The hardware subsystems include the design of and build of a mechanical motor mount and housing using CAD software and 3d printing. It also includes a component selection process which aims to select the best possible components that meet the user requirements and maximizes performance whilst still remaining within budget constrains. The Raspberry Pi 3B was chosen as the embedded system for this project along with the Adafruit AMG8833 IR Thermal Camera and the 5MP Raspberry Pi visual camera. The motor chosen for this project is the Fitec micro 9g servo motor as it is easy to use and power efficient whilst also being non-expensive.

The software subsystems include the design of an algorithm to detect humans objects on a visual frame captured by the Raspberry Pi visual camera and actuate the servo motor to centralise the object on the frame. It also includes the design of an algorithm which makes use of sensor fusion to combine the data from the visual and thermal camera to improve detection and object tracking of the system under poor visual conditions. OpenCV and Python along with a custom trained YOLO-V4 Tiny object detector were the main tools used to implement both algorithms.

To ensure the integrity of each subsystem various acceptance test protocols(ATPs) were laid out for each subsystem. These ATP's were specifically designed to test the performance of each subsystem individually.

3.2 User Requirements

The following user requirements were derived from the problem statement, scope breakdown and project objectives in Chapter 1.

Table 3.1: User requirements of Camera Tracking System

Requirement ID	Requirement Text
RU001	The system must be able to track at least a single object in a visual scene
RU002	The system must incorporate a low-cost thermal camera to improve tracking under poor visual conditions
RU003	The system must be mounted on a mechanical platform to move the cameras and keep the target within the visual frame with at least one degree of freedom.

3.3 User Requirement Analysis

This section aims to refine the user requirements into functional requirements.

3.3.1 Analysis of RU001

"The system must be able to track at least a single object in a visual scene"

The system must use an object detection algorithm to identify objects on a visual scene and actuate a motor to track the position of the object.

Req. ID	**Requirement Text**	**Derived From**
FR001	Development of an object detection algorithm to detect a human in a visual screen	RU001
FR002	Detection algorithm should be able to actuate a motor to track the position of an object on a visual camera	RU001
FR003	The system must include a visual camera which can be used to capture visual scenes	RU001
FR004	The system must include an embedded system which can be used to run the algorithms, interact with the camera and actuate the motor	RU001

Verification of FR001
The object detection algorithm will be set to detect only humans on a visual scene. This will be tested by running the algorithm with a visual camera and checking if it draws a boundary box around people on the frame whilst ignoring other objects.

Verification of FR002
Place an object off centre on the visual frame and if the system works as intended the algorithm should actuate the motor to turn the camera such as to centralise the object on the frame.

3.3.2 Analysis of RU002

"The system must incorporate a thermal camera to improve tracking under poor visual conditions"

The system must make use of an algorithm that incorporates sensor fusion techniques to improve the tracking performance of the visual camera under poor visual conditions.

Req. ID	Requirement Text	Derived From
FR005	Development of an algorithm that uses both visual and low cost thermal camera data to improve performance under poor visual conditions	RU002
FR006	The system must include a low cost thermal camera to capture infrared scenes in the environment visual conditions	RU002

Verification of FR005

Run the system under varying visual conditions with just the system with only the visual camera (Thermal camera disabled) and then with the thermal camera enabled (Fused System). If the sensor fusion algorithm works as intended there should be an improvement in object detection under poor visual conditions (i.e Dark environments) with the thermal camera enabled.

3.3.3 Analysis of RU003

"The system must be mounted on a mechanical platform to move the cameras and keep the target within the visual frame"

The hardware design must include a housing for a the motor and any necessary electronics needed for the system. It must also include a mounting platform to mount both the thermal and visual cameras on the motor.

Req. ID	Requirement Text	Derived From
FR007	Design and building of a mechanical housing to mount electronics and motor	RU003
FR008	Design and building of a mechanical mount to mount both the visual and thermal camera onto the motor. The cameras must rotate with one degree of freedom in the x-axis only, thus meaning they are stationary in the y and z planes respectively	RU003
FR009	The Motor should be able to handle the weight of both cameras as well as the weight of the mechanical motor mount	RU003

Verification of FR007

The housing will be designed with enough space to fit all the necessary components with dedicated mounting slots for the motor and embedded device.

Verification of FR008

The mechanical motor mount will be designed to house both cameras and the motor will be mounted on the housing such that it rotates within the x-plane.

3.4 Subsystem Identification

Based on the functional requirements set out above the camera tracking system can be divided into various subsystems which will be designed and tested individually to ensure the optimal performance of the final design. Figure 3.2 below shows how the overall system is sub-devised into subsystems.

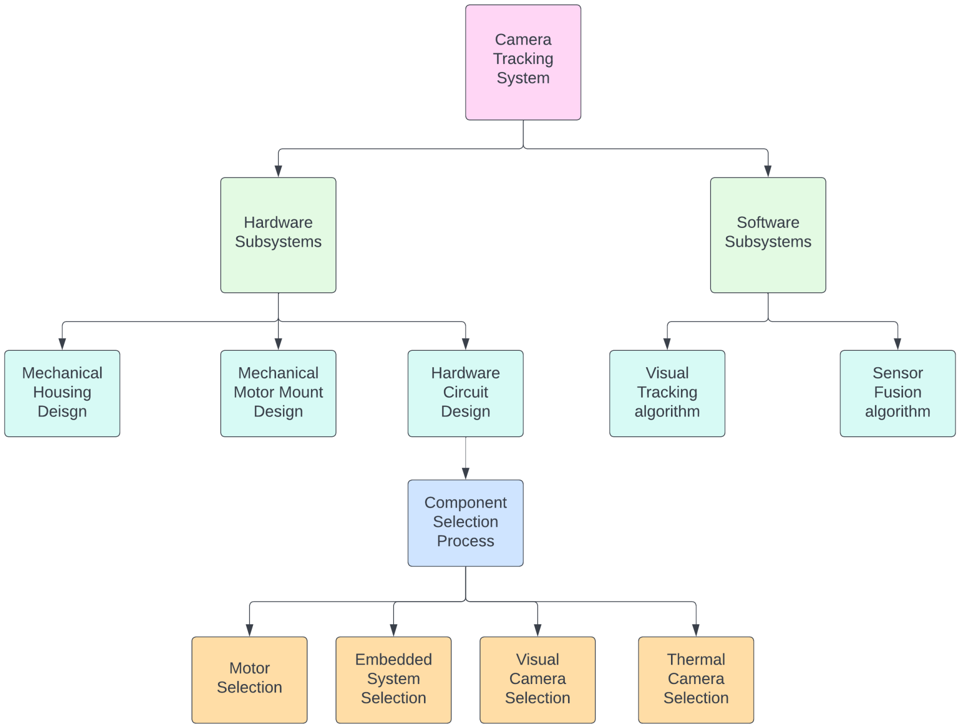


Figure 3.2: Subsystem Breakdown of the Camera Tracking System

Chapter 4

Design

This chapter aims to give a detailed account of the design process for the camera tracking system. This includes a full technical description of both the hardware- and software-subsystem design along with suitable acceptance test protocols for each subsystem.

4.1 Hardware design

As seen in section 3.4 the hardware component of the camera tracking system can be broken up into various subsystems. This section covers the detailed design process of each hardware subsystem whilst also proposing suitable acceptance test procedures which aim to evaluate the subsystems performance in reference to the user demands. As this project has budgetary constraints the main aim of this section is thus to produce suitable hardware solutions which deliver the best possible results within the given budget.

4.1.1 Component Selection

As shown in figures 3.2 and 4.1 various components are needed in the camera tracking system which all serve specialised roles in fulfilling specific functional requirements. As mentioned in section 1.5 this project has a budget constraint of R2000 and as such a rigorous component selection process is required to select the best components for the price. The selection criteria used for component selection can be seen below in order of importance:

- **Performance Quality** - To ensure the best possible performance of the camera tracking system it is preferential to have the best components on the market.
- **Price** - Due to budget constraints a price-performance trade off is needed

- **Compatibility** - It is important to ensure that the components are compatible with each other and that they use similar communication protocols
- **Availability** - As the project has a strict deadline it is important to select components that are readily available
- **Flexibility** - Due to budget constraints it would be preferential if one component can fulfil multiple system requirements

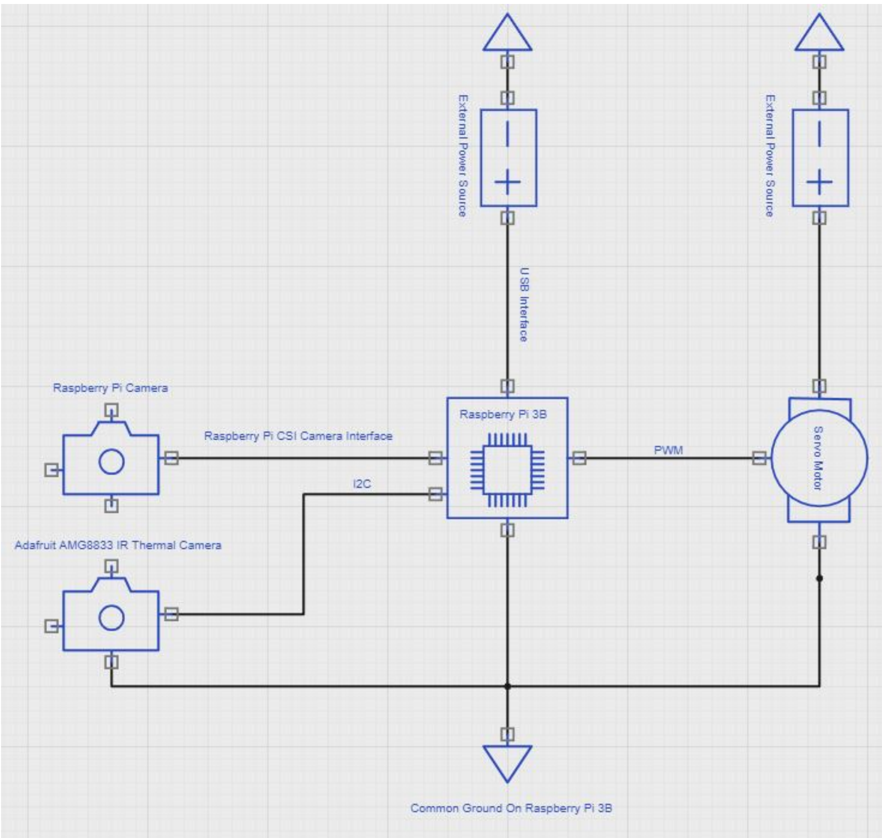


Figure 4.1: High level diagram showing how the hardware components are connected and their respective communication protocols

Raspberry-Pi 3B

The Raspberry Pi 3B is a powerful micro-controller device with a Quad Core 1.2GHz Broadcom BCM2837 64bit CPU and 1GB Ram. It also boasts 28 general input-output (GPIO) pins including various pins with support for PWM and multiple communication protocols including UART, I2C, and SPI build in. It also has support for serial communication over micro USB with the micro USB port also serving as the 5V power-in port. The dedicated CSI camera port on the raspberry pi 3B allows for the easy connection of a raspberry pi visual camera and eliminates the need for any unnecessary overhead circuitry.

Due to the fact that the raspberry pi 3B provides moderately powerful performance, supports multiple communication protocols and PWM for servo motor control, has a dedicated CSI camera port and was readily available at no cost it was chosen as the embedded system for this project.

Other devices Considered
The Raspberry Pi 4 and the NVIDIA Jetson Nano were also considered as possible embedded system candidates as they are more powerful than the Raspberry Pi 3B. However the Raspberry Pi 4 is unavailable on the market and the Jetson Nano is too expensive.

Figure 4.2: Figure depicting the Raspberry Pi 3B [10]

Raspberry Pi 5MP(1080P) Mini Camera-Video

The Raspberry Pi camera boasts a 5MP camera lens with 1080p video capability in a small compact design of 25mm x 24mm x 9mm and a weight of just over 3 grams. It communicates with the Raspberry Pi device through a dedicated CSI camera interface and comes pre-made with 4 mounting holes.

Due to its compact and light design, 1080p video capabilities along with the ease of communication with a Raspberry Pi 3B device through CSI the Raspberry Pi camera Mini is the perfect visual camera for this project. It is also very inexpensive with a price of only R165.00.

Other devices Considered The Raspberry Pi Camera Board, Version 2, Sony IMX219 8-Megapixel device was also considered for this project. It has a better camera module (8MP) which allows it to give clearer images than the camera selected for this project. However it wasn't selected as it is much more expensive and the increase in video quality has a negligible effect on the performance of the overall system.

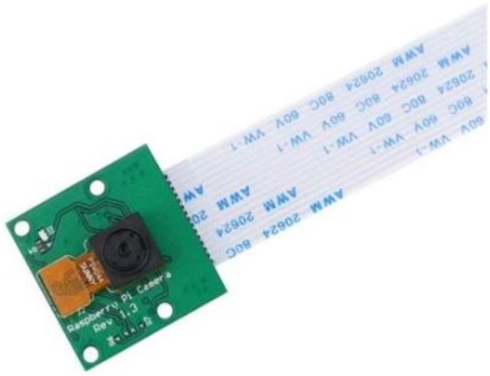


Figure 4.3: Figure depicting Raspberry Pi 5MP(1080P) Mini Camera-Video[11]

Adafruit AMG8833 IR Thermal Camera

The Adafruit AMG8833 is a 8x8 thermal camera with the capability of measuring temperatures ranging from 0° to 80° with an accuracy of ±2.5° . It weighs approximately 6 grams with a compact board design of 25.8mm x 25.5mm x 6.0mm. The camera has a maximum refresh rate of 10Hz and communicates with embedded devices over the I2C communication interface. This camera module has price of R867.00.

The low resolution of 8x8 coupled with the slow refresh rate 10Hz means that the Adafruit AMG8833 takes up minimal computational resources for image processing applications. This important as the Raspberry Pi 3B has limited computational resources. Furthermore the Adafruit AMG8833 also satisfies RU002 which states that the thermal camera must be of low-cost. Therefore it was chosen as the thermal camera for this project.

Figure 4.4: Figure depicting the Adafruit AMG8833 device[12]

Feetech Micro 9g Servo Motor

The Feetech Micro 9g is a small servo motor that can produce a maximum torque of 1.3 kg · cm @ 4.8V, 600mA. It can reach a maximum rotational speed of 0.12 sec/60°

and has a a running angle of approximately 120° . The position of the motor can be controlled by providing it with a PWM signal with a pulse with between 900 s(Minimum angle) and 2100 s (Maximum angle) over the control lead. The Feetech Micro 9g is very inexpensive with each unit pricing only R68.00.

Due to the lightweight designs of both the visual and the thermal camera the motor needed for this project doesn't need to have a lot of torque to perform adequately. Thus the Feetech Micro 9g is perfect as it is easy to control and very inexpensive.

Detailed Wiring Diagram for Hardware Components

Figure 4.5 below shows a detailed wiring diagram for connecting the Adafruit AMG8833 IR Thermal camera and the servo motor to the Raspberry Pi GPIO pins. The Raspberry Pi Camera is not included on the wiring diagram as it connects directly to the CSI connector in the Raspberry Pi 3B device.

The power required for the servo motor can be provided by any external power supply that can supply at least 750mA at 5V DC. The Raspberry Pi device can be powered over the micro USB interface and requires a power supply that can provide a maximum output of at least 1.3A at 5V DC.

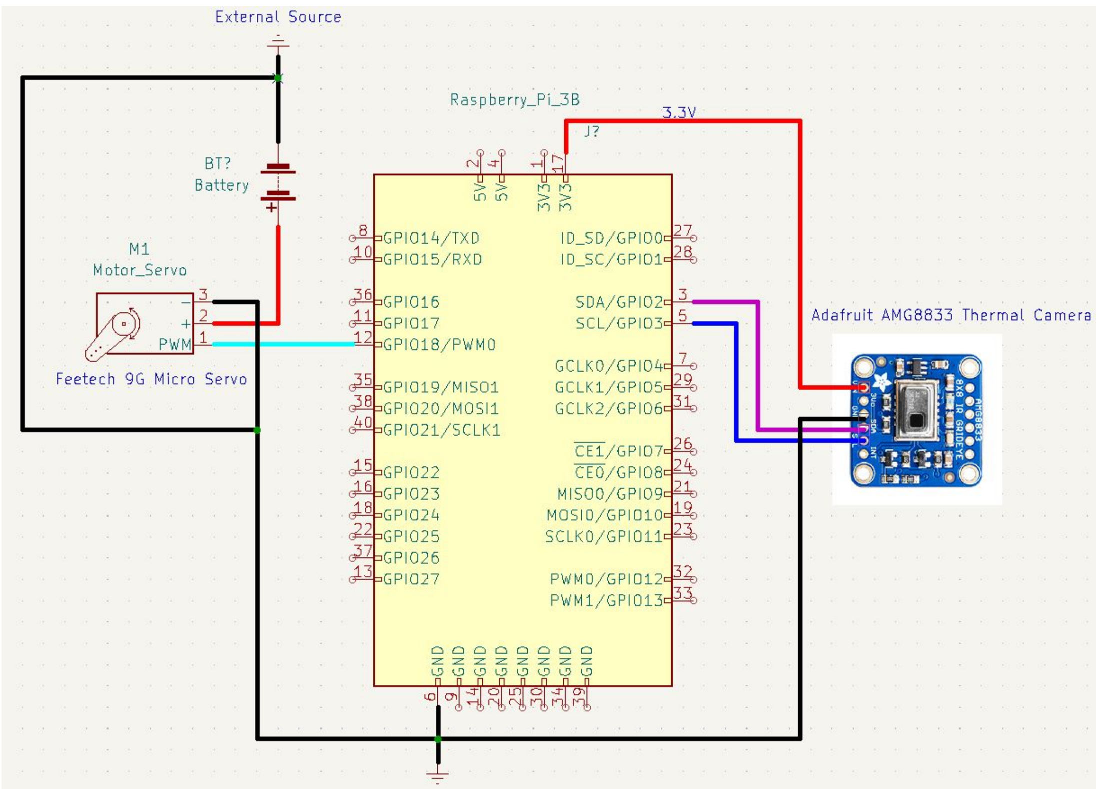


Figure 4.5: Detailed wiring diagram of hardware components for the camera tracking system

4.1.2 Mechanical Housing Design

As seen in Section 3.3 functional requirement FR007 states that the mechanical housing must have enough space to fit all electronic components and possess dedicated mounting slots for the Raspberry Pi 3B and the Feetech Micro 9g servo Motor. To fulfil all the necessary requirements this project makes use of a three tier stack design with the first tier housing the Raspberry Pi 3B, The second tier acting as a protective layer for the Raspberry pi and the third tier housing the Feetech Micro 9g servo motor. All three tiers were designed using the TinkerCad software and manufactured using the Prusca Mini 3D printer.

The first tier if the housing is specifically designed to mount the Raspberry Pi 3B device, thus it has 4 mounting points which align directly with its dedicated mounting holes. The first tier also aims to provide extra space for any electronics that might be added to the design later (i.e, External Power Supply) therefore the design incorporates extra height elevation under the Raspberry Pi Mounting points. To encompass all the above set parameters the first tier has the following external dimensions: 89.92mm x 61.72mm x 39.14mm.

The design file for the 2nd tier of the mechanical housing can be found on gitlab at the following link: *Base Final Design*

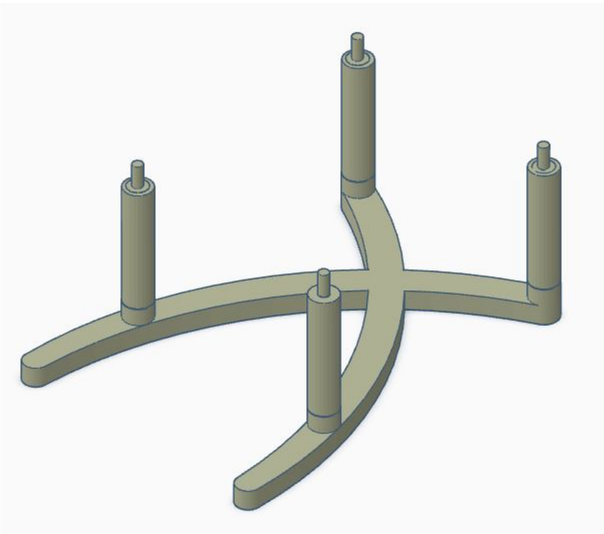


Figure 4.6: Figure depicting the 1st tier of the mechanical housing design

The second tier of the mechanical housing is simply designed to improve the aesthetic of the housing and to provide a protective cover for the Raspberry Pi. It has four mounting holes which align perfectly align with the first tier Raspberry Pi mounting holes. Therefore the second tier of the housing has the following external dimensions: 64.18mm x 54.19mm x 16.10mm. The design file for the 2nd tier of the mechanical housing can be found on gitlab at the following link: *Middle final Deign*

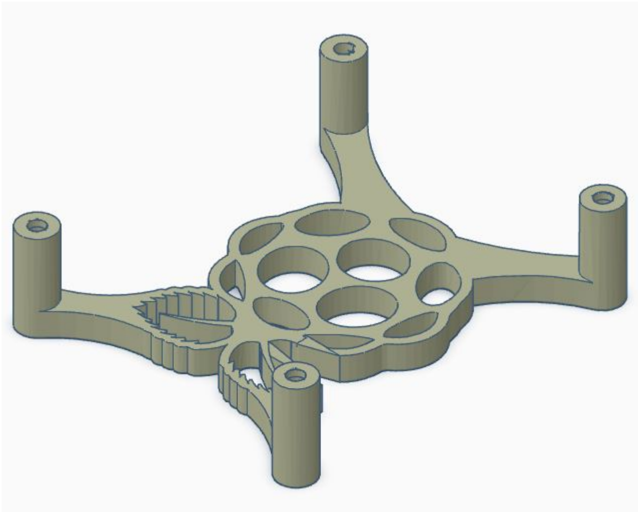


Figure 4.7: Figure depicting the 2nd tier of the mechanical housing design

The 3rd tier of the mechanical housing is designed to house the Feetech Micro 9g servo motor. It has 4 mounting points that perfectly align with the mounting holes of the second tier. Thus it has the same external length and width as the second tier with the height tailored to fit the servo motor specifically. As a result the 3rd tier has the following external dimensions: 64.18mm x 54.19mm x 55.09mm. The design file for the 3rd tier of the mechanical housing can be found on gitlab at the following link: *Top Final Design*

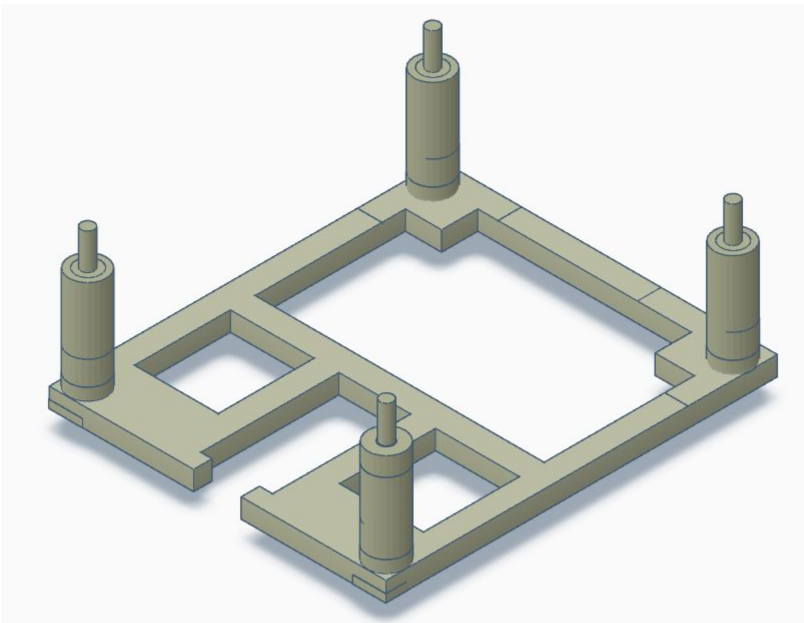


Figure 4.8: Figure depicting the 3rd tier of the mechanical housing design

4.1.3 Mechanical Motor Mount

The mechanical motor mount for this system must be able to house both the thermal and the visual camera whilst also being lightweight to ensure that the servo motor can still turn freely. Given these specifications the following design was produced.

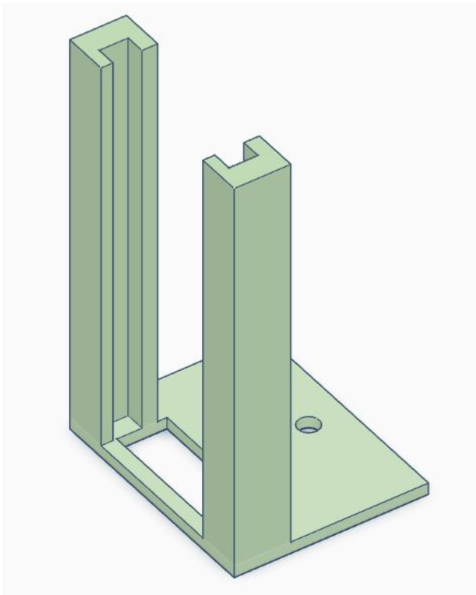


Figure 4.9: Figure depicting the mechanical motor mount

The design in figure 4.9 can be easily attached to the servo motor by fitting a screw through the mounting hole in the middle. The Raspberry Pi- and thermal camera can both slide into the vertical slit in any order, however the preferred configuration would be to place the Pi-camera at the bottom. The design makes provision for the raspberry pi ribbon cable to run out the bottom of the mount to help prevent any damage to the cable due to sharp bending. The external dimensions of the motor mount are as follows: 26mm x 32mm x 55mm with a ⌀3mm mounting hole. The design file for the mechanical motor mount can be found on gitlab at the following link: *Motor Mount Design*

4.1.4 Acceptance Test Protocols: Hardware Design

Table 4.1: Hardware subsystem acceptance test protocols

ATP ID	**Hardware Subsystem**	**Acceptance Test Protocols(ATP)**	**Functional Requirement Satisfied**
ATP1	Raspberry Pi 3B	The device turns on correctly and the LED turns on when powered on. Device can be accessed successfully over ssh	FR004
ATP2	Raspberry Pi 5MP(1080P) Mini Camera-Video	LED turns on when device is powered on and no errors are encountered when running following line in the raspberry pi command line: $ raspistill -o test.jpg	FR003
ATP3	Adafruit AMG8833 IR Thermal Camera	Device is detected over I2C when plugged in. Device should output a 2d array of temprature values when running the following script: *Adafruit AMG833 Test*	FR006
ATP4	Feetech Micro 9g Servo Motor	Device should move to three different positions when running the following script: *Motor Test*. Device should be able to turn effectively with cameras and mount attached	FR002, FR009
ATP5	Mechanical Housing	The housing should have dedicated mounting positions for the servo motor and the raspberry pi device. Housing should have space to potentially fit in other electronics if needed.	FR007
ATP6	Mechanical Motor Mount	Mechanical mount should be able to hold both the thermal and the visual camera. Mount should fit onto the Feetech 9g Servo motor	FR008

4.2 Software Design

As seen on figure 3.2 the software portion of this project can be divided into two sub-categories. The first sub-category aims to develop an algorithm to track a human object on a visual scene captured by the raspberry pi camera. The second sub-category aims to improve on the visual tracking algorithm by making use of sensor fusion to incorporate a low-cost thermal camera into the design.

4.2.1 General Design Choices

The software for this project was designed using the Python programming language as it has extensive support for computer vision applications through the OpenCV library. It also allows for the easy control of servo motors with minimal motor jitter by using the pigpio library.

To perform human-object detection on visual frames this project makes use of the YoloV4-tiny object detection algorithm. As seen on figure 2.4 it has the fastest inference time compared to other well known object detection algorithms whilst also providing moderately good detection accuracy. The higher inference time is especially important as it allows for faster performance on devices that have very limited computational resources such as the Raspberry Pi.

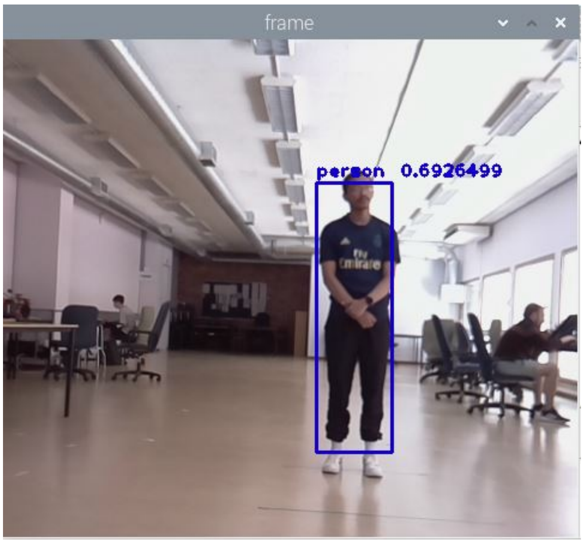


Figure 4.10: Example of a human-object detected by the Yolov4-tiny on a visual frame captured by the Raspberry Pi camera

The poor 8x8 resolution of the AMG8833 Thermal camera doesn't allow for the use of a CNN based object detection model such as Yolo for human-object detection. As a result this project makes use of the background subtraction model to perform this task.

After testing the AMG8833 thermal camera it was found that human-objects show up as white spots on the RGB frame as shown in figure 4.11

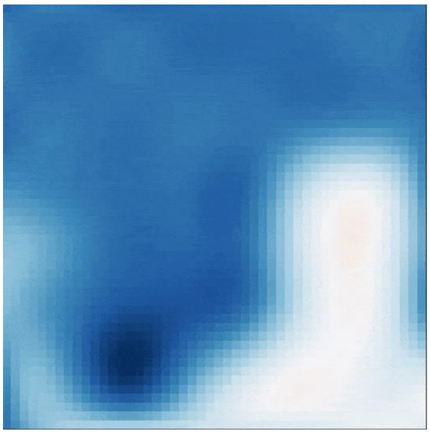
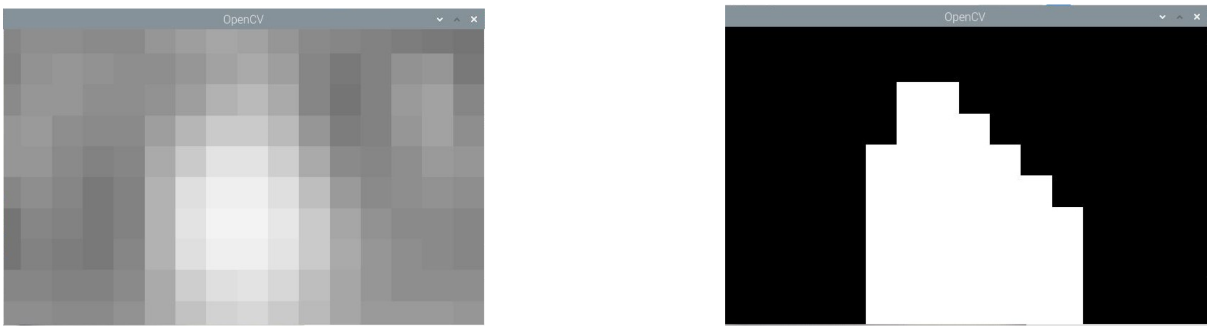


Figure 4.11: Figure of a human-object captured by the AMG8833 Thermal Camera in the RGB Color space [13]

Thus the background subtraction model works by converting the original thermal image to gray-scale and then applying a binary mask that turns all the pixels that are non-white completely black and all the white pixels completely white as seen in figure 4.12 below.



(a) Original thermal image with gray-scale filter

(b) Thermal Image after Binary mask is applied

Figure 4.12: Comparison between gray-scaled thermal image containing a human object before and after binary mask is applied

After the binary mask is applied the model looks for the white spot with the biggest pixel area as it is most likely the human-object in the frame. Similarly any white-object with pixel area less than 50 pixels is ignored by the model as these are very small objects and most likely not a human. However the model will not be able to distinguish a human object from any other object white object with a pixel area bigger than 50. This could possibly lead to false detections if foreign objects enter the field of view of the camera.

4.2.2 Visual Camera Tracking Algorithm

The diagram below shows a high level logic diagram of the tracking algorithm using the visual camera only.

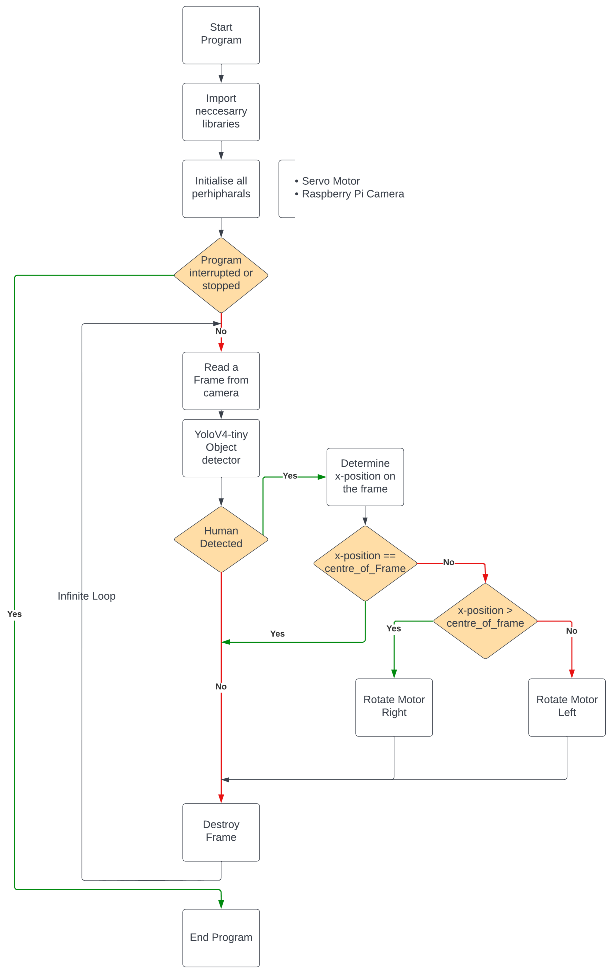


Figure 4.13: Logic flow diagram of the visual tracking system algorithm

The algorithm shown in figure 4.13 above follows a simple logic sequence. It works by reading in a frame from the Raspberry Pi camera and feeding it into the YoloV4-tiny object detector. If a person is detected the algorithm will determine the x-position of the person on the frame in terms of the pixel value. If the person is in the centre of the frame the loop will simply move on to the next iteration, however if the person is not in the centre the algorithm will actuate the motor to turn the camera left or right given the x-position of the person relative to the centre of the frame.

The detailed code for this algorithm can be found at the following git repository link: *Visual Camera Tracking Algorithm*

4.2.3 Visual-and Thermal-camera fusion Algorithm

The diagram below shows a high level logic flow diagram of the tracking algorithm using the visual and thermal camera.

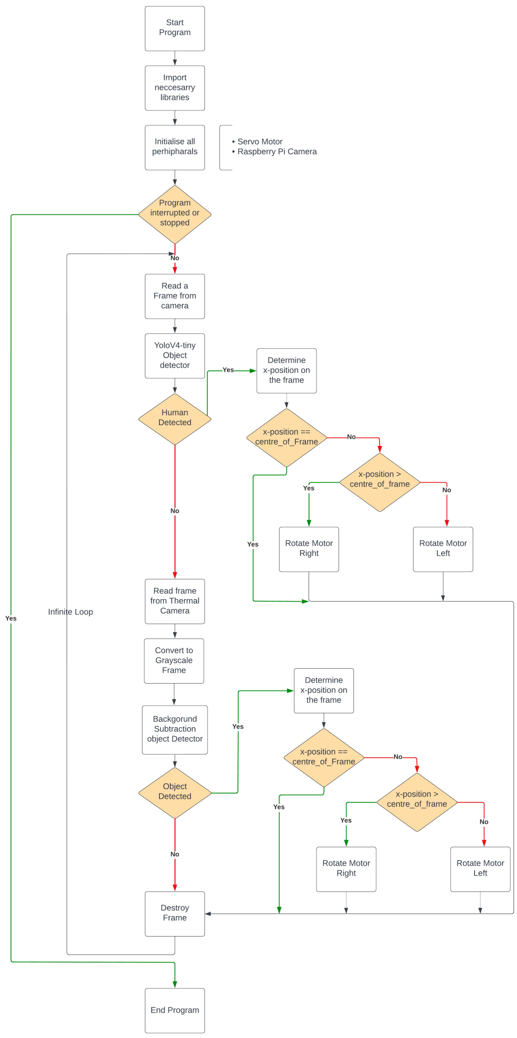


Figure 4.14: Logic flow diagram of Visual-and Thermal-camera fusion Algorithm

As seen from figure 4.14 the visual camera acts as the primary source for tracking in the fusion algorithm with the thermal camera only being used when no human-object is detected on the visual frame (i.e, in poor visual conditions). This is as due to the thermal camera having a much higher chance of false detections than the YoloV4-tiny detection on the visual camera. When a human-object is detected on the visual frame the algorithm completely ignores the thermal camera and immediately jumps to the end of the main loop. This is done to help reduce the amount of computational resources used and speed up the tracking on the visual camera when the thermal camera is not being used for tracking. The two sensors work independently to achieve tracking under varying light conditions, thus the system makes use of the complementary sensor fusion technique shown of figure 2.8.

The detailed code for the fusion algorithm can be found at the following git repository link: Visual-and Thermal-camera fusion Algorithm

4.2.4 Acceptance Test Protocols: Software Design

Table 4.2: Software subsystem acceptance test protocols

ATP ID	Software Subsystem	Acceptance Test Protocols(ATP)	Functional Requirement Satisfied
ATP7	Visual Camera Tracking Algorithm	• Algorithm compiles and runs successfully without errors • Algorithm is able to detect a human-object on a visual frame • Algorithm is able to actuate the servo motor to track a person on a visual frame	FR001,FR002
ATP8	Visual-and Thermal-camera fusion Algorithm	• Algorithm compiles and runs successfully without errors • Algorithm is able to detect a human-object on a visual frame • Algorithm is able to detect at least one human object on a thermal frame in poor visual conditions • Algorithm is able to actuate the servo motor to track object on a visual frame • Algorithm is able to actuate the servo motor to track an object on a thermal frame	FR001,FR002, FR005

4.3 Design Implementation

The physical construction of the final prototype for the system will be explored in this section along eith the challenges faced during the implementation of the design.

The system followed to wiring diagram shown on figure 4.5 and made use of the female-to-female and Male-to-female pin connectors to connect the Adafruit AMG8833 Thermal camera and the servo motor to the Raspberry Pi GPIO pins. The Raspberry Pi camera was then connected through the CSI camera port on the Raspberry pi device. However the ribbon cable that came included with Raspberry Pi camera was to stiff and couldn't bend enough to fit into the mechanical housing. As a result the casing had to be re-design and made bigger to accommodate the cable.

All the parts of the mechanical housing and motor mount were printed using the Prusa Mini 3d-printer. This was an iterative process as well due to some of the 3d printed parts being very fragile and breaking easily. The legs used to interlock each part of the mechanical housing were especially fragile as they on the original design the were only 2mm in diameter and hollow with an inner diameter of 1mm. After a few iterations the diameter of the legs were increased to 3mm with a solid infill and this made them a lot more durable.

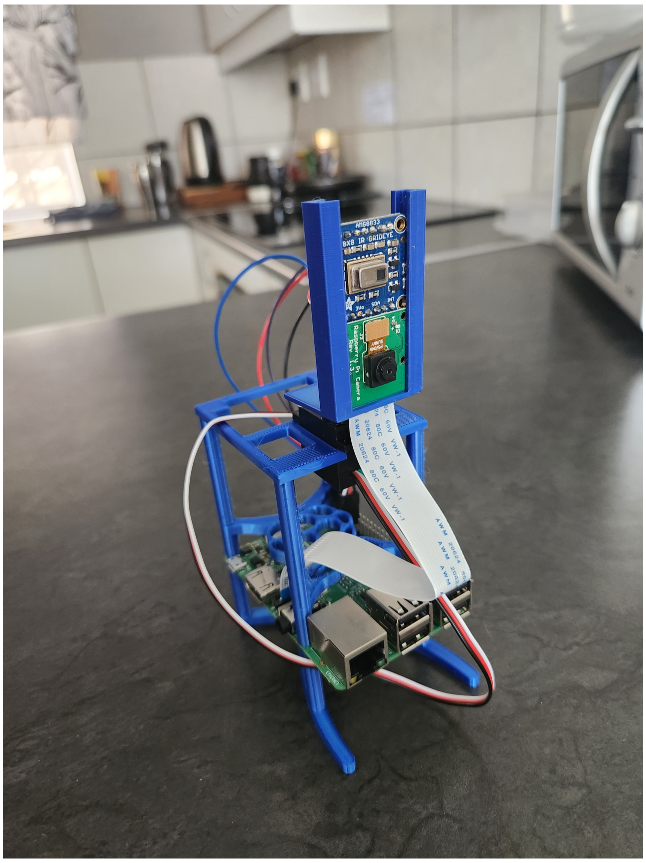


Figure 4.15: image of the final implementation of the device

4.4 Acceptance Test Protocol Results

During the design implementation phase of the project all the hardware ATP's shown in Table 4.1 were tested. The table below shows the outcome of each ATP.

Table 4.3: Results of hardware ATP's

ATP ID	**Result**	**Justification**
ATP 1	Pass	The Raspberry Pi device worked as intended and could be accessed over ssh through Virtual Studio.
ATP2	Pass	The Raspberry pi camera was able to capture visual frames from the environment and feed them into the tracking algorithm.
ATP3	Pass	The thermal camera was able to capture thermal data from the environment and all the data could be accessed over the I2C interface on the Raspberry Pi.
ATP4	Pass	The servo motor responded to PWM commands send from the Raspberry Pi through GPIO Pin 12. The motor was also able to turn effectively with the cameras mounted on top.
ATP5	Pass	The housing was able successfully hold both the Raspberry Pi device and the servo motor as seen in figure 4.15.
ATP6	Pass	The mount fit onto the servo motor and had enough space to fit both the visual and the thermal camera as seen in 4.15.

The software ATP's we tested in the results chapter of the thesis as part of the performance of the system.

Chapter 5

Results

The main aim of this chapter is to provide an insight into the performance of the camera tracking system in relation to certain performance metrics. It will start off by looking at the experimental setup used for testing the system and explain why certain testing parameters were chosen. Thereafter it will move onto the performance metrics used to evaluate the systems performance. Finally it will present and analyse the findings of the experiments.

5.1 Experimental setup

To ensure the accuracy of the test results obtained during the testing phase the device will be tested in two different environments. This will help provide a deeper insight into the performance of the device under varying conditions. The first environment is the CRG Lab at UCT. This environment is a big open space with little to no external factors that could possibly generate noise on the cameras.

The second testing environment is a living room/Kitchen of a residential house as there are a lot of devices such as computers, kettles, space heaters as well as other humans that generate heat which could serve as possible sources of noise for the thermal camera. The idea behind testing the device in this environment is to evaluate its performance in a more realistic setting.

The device will also be placed at a height of approximately 1m off the ground in all testing scenarios to allow any human that enters the frame to be in the field of view of both cameras.

5.2 Performance Metrics

As the two camera sensors work independently in the fusion algorithm the overall performance of the system can be tested by evaluating the performance of each camera module individually. As a result the following metrics will be tested to evaluate the performance of the system:

- Noise rejection of the system
- Detection range of the visual camera tracking system
- Performance of the visual tracking algorithm under varying light conditions
- Detection range of the thermal camera object detection algorithm
- Tracking Speed of fusion tracking algorithm

5.3 Experimental Results

This section aims to present the findings of the experiments done to test the performance of the overall system in reference to the performance metrics set out above. It will also outline the testing procedure used to test each metric as well as provide a brief discussion of the results obtained.

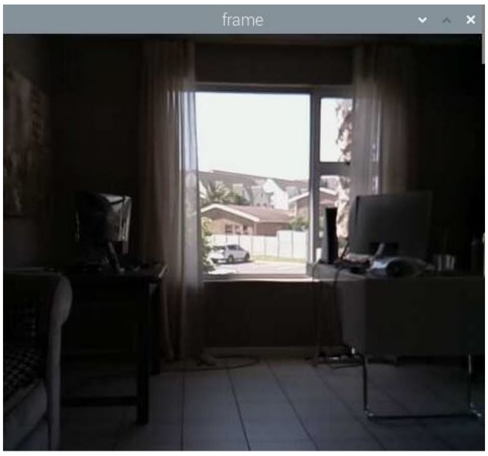
To better recognise the detected objects/noise on the thermal frames all the results below display the binary mask developed from each thermal frame rather than the raw unprocessed frame data. The differences between the raw frame and the image after the binary mask is applied can clearly be seen on figure 4.12. The figure with the binary mask applied clearly indicates the position of the detected object better than the raw gray-scaled image which is filled with clutter.

5.3.1 Noise rejection of the system

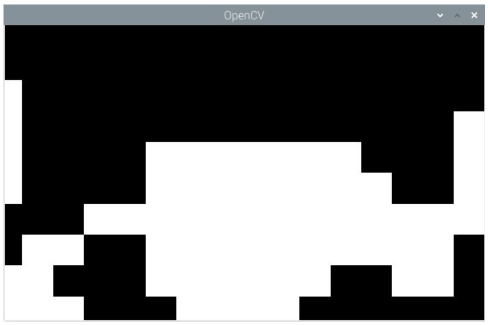
Testing Procedure

To test the noise rejection capabilities of the system to environmental elements the device was placed in an empty space without any human-objects in the field of view of the cameras. This was be done in both the CRG Lab and Living Room environments.

Experiment Results and Discussion

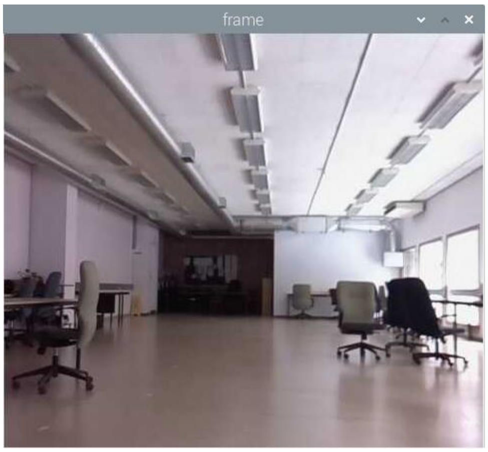


(a) Living Room area captured with visual camera at optimal lighting conditions with no human-object in the frame

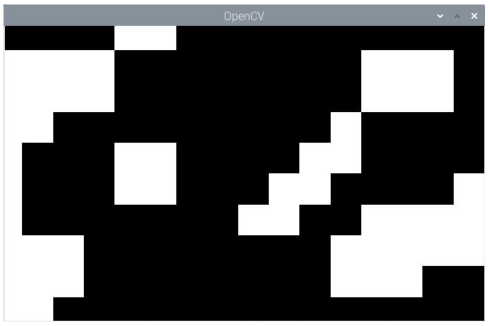


(b) Living Room area captured with thermal camera with no human-object in the frame

Figure 5.1: The two figures above show the images captured by the visual and thermal camera of the living room environment at optimal lighting with no human-object on the frame



(a) CRG Lab area Captured with the visual camera at optimal lighting conditions with no human-object in the frame



(b) CRG Lab area Captured with thermal camera with no human-object in the frame

Figure 5.2: The two figures above show the images captured by the visual and thermal camera of the CRG Lab environment at optimal lighting with no human-object on the frame

Figures 5.2 and 5.1 showcase the images taken at the CRG Lab and the living room environments in optimal lighting conditions without a human objects in the field of view of the cameras. This was done to test the effect of the environment on both detection algorithms in order to see if there are any false detections.

From both 5.2.a and 5.1.a we can see that the visual camera performs as expected and no false detections were made. However the thermal camera was heavily effected by environmental noise as seen from figures 5.2.b and 5.1.b.

The noise on the thermal camera is most likely due to the camera picking up traces of thermal radiation from the direct sunlight hitting the camera lens. This can especially be seen in figure 5.1.b where the there is a large amount noise detected directly in the middle of the thermal frame which correlates directly to the window on the related visual frame.

5.3.2 Detection range of the visual camera tracking system

Testing Procedure

To test the detection range the system the visual tracking algorithm was ran with a human-object directly in the centre of the visual frame. The person started off directly in front of the camera and slowly moved backwards until they were outside the range of detection. The maximum distance of detection was then be measured by taking the last point where the object detection algorithm recognised the person on the frame.

Experimental Results and Discussion

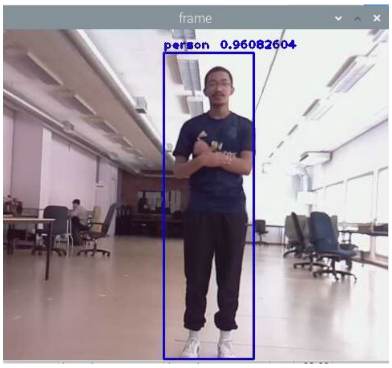


Figure 5.3: Image showcasing a human-object detected on a visual frame captured by the Raspberry Pi camera at a distance of 2m in optimal lighting conditions

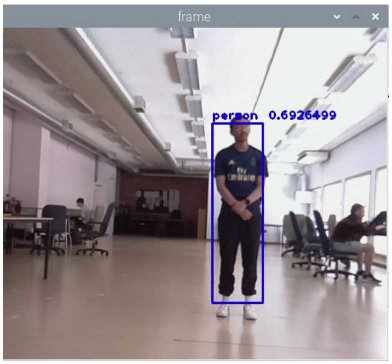


Figure 5.4: Image showcasing a human-object detected on a visual frame captured by the Raspberry Pi camera at a distance of 5m in optimal lighting conditions

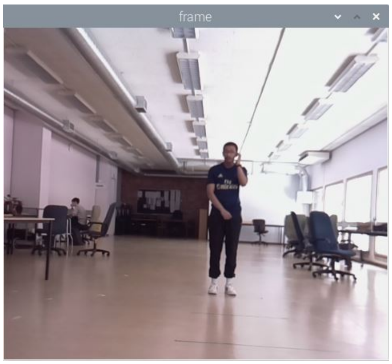


Figure 5.5: Image showcasing a human-object at a distance of 6m on a visual frame captured by the Raspberry Pi camera in optimal lighting conditions

Figures 5.3, 5.4 and 5.5 above showcase the performance of the visual tracking algorithms ability to detect human-objects over a varying distances in optimal lighting conditions. The figures also include the confidence level of the YoloV4-tracking algorithm which is an indication of how certain the model is of its prediction.

From figures 5.3 and 5.4 we can clearly see that the model makes suitable predictions up to 5m. However the confidence level of the prediction drops from 0.961 to 0.692 when the person moved from 2m to 5m away from the camera. This indicates that the model becomes less accurate the further away a target is.

From figure 5.5 we can see that the model cant detect a human object that is at 6 meters away from the camera, thus meaning that the maximum detection range of the visual tracking algorithm is approximately 5m.

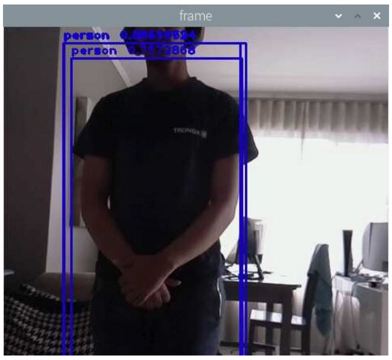


Figure 5.6: Image showcasing a human-object detected on a visual frame captured by the Raspberry Pi camera at a distance of 1m in optimal lighting conditions

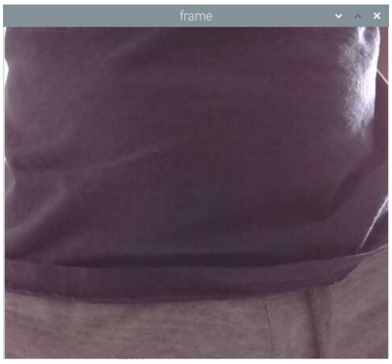


Figure 5.7: Image showcasing a human-object captured by the Raspberry Pi camera at a distance of 0.3m in optimal lighting conditions

From figures 5.7 we can see that if the target is closer than 1 meter away from the camera the face of the person is no longer in the frame and as a result the detection algorithm cant detect the person on the frame. If the target is a exactly 1 meter away from the camera the person can be detected moderately well as seen on figure 5.6. Thus meaning a person must be at least at least 1 meter away from the camera in order for the device to be able to track them.

5.3.3 Performance of the visual tracking algorithm under varying light conditions

Testing procedure

To test the performance of the visual tracking algorithm under varying light conditions the device was tested at different times of the day. In each test case a human-object was placed at 2m away from the camera to ensure that the distance of the person away from the camera didn't effect the confidence level of the detection algorithm and that the only variable was the change in lighting.

Experiment Results and Discussion

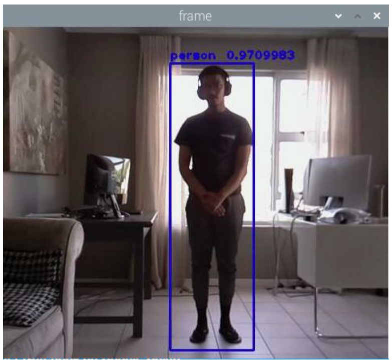


Figure 5.8: Image showcasing a human-object detected on a visual frame captured by the Raspberry Pi camera at a distance of 2m in optimal lighting conditions - Living Room environment

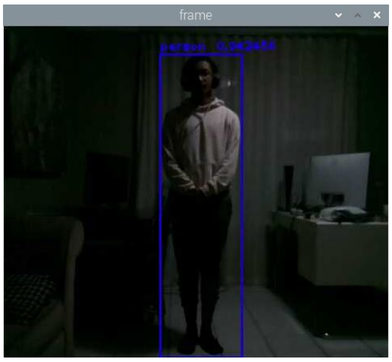


Figure 5.9: Image showcasing a human-object detected on a visual frame captured by the Raspberry Pi camera at a distance of 2m in dim lighting conditions - Living Room Environment

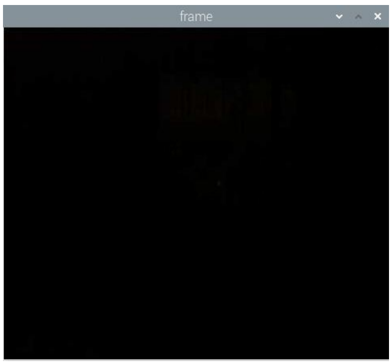


Figure 5.10: Image showcasing a human-object detected on a visual frame captured by the Raspberry Pi camera at a distance of 2m in a dark room - Living Room

Figures 5.8, 5.9 and 5.10 above showcase the performance of the visual tracking algorithms performance under varying light condition in the same environment.

From figures 5.8 and 5.10 it is clear that the camera performs as expected in optimal- and dark-lighting conditions with the model having a confidence level of 0.971 optimal lighting conditions while no person is detected in the dark environment. Surprisingly, the model also performed relatively well in dim lighting conditions with a detection confidence level of 0.943 as seen on figure 5.9.

These results show that the model can perform relatively good human detection on the visual camera within 2 meters under any lighting circumstances except in environments with very little to no light.

5.3.4 Detection range of the thermal camera object detection algorithm

Testing Procedure

To test the detection range of the thermal camera object detection algorithm the device will be placed in a darkroom with a human-object directly in the centre of the thermal cameras field of view. The person will start close to the camera and slowly move backwards until they are outside the detection range of the algorithm

Experiment Results and Discussion

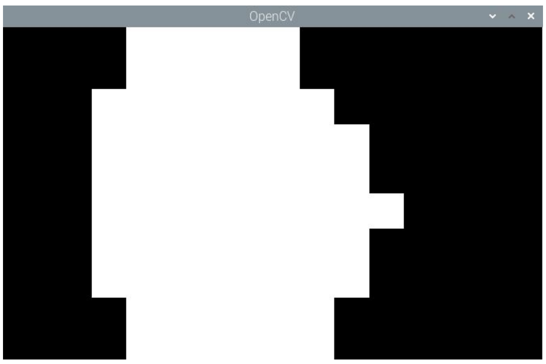


Figure 5.11: Image showcasing a human-object detected on a thermal frame captured by the Adafruit AMG8833 camera with at a distance of 1m in a dark room

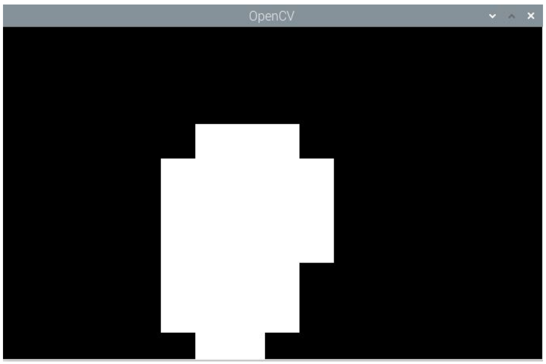


Figure 5.12: Image showcasing a human-object detected on a thermal frame captured by the Adafruit AMG8833 camera with at a distance of 2m in a dark room

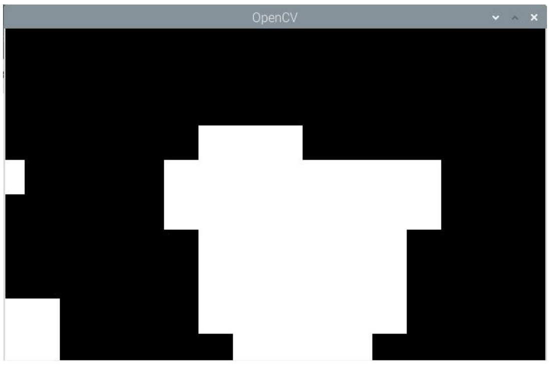


Figure 5.13: Image showcasing a human-object detected on a thermal frame captured by the Adafruit AMG8833 camera with at a distance of 3m in a dark room

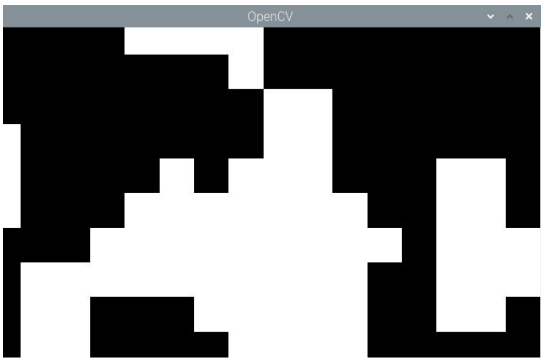


Figure 5.14: Image showcasing a human-object detected on a thermal frame captured by the Adafruit AMG8833 camera with at a distance of 4m in a dark room

Figures 5.11, 5.12, 5.13 and 5.14 above showcase the performance of the thermal camera detection algorithm over various distances in a dark environment.

From figures 5.11 and 5.12 it is clear that the background subtraction algorithm detects a person perfectly on the thermal frame within 2 meters from the camera which means that the tracking within this range functions really well.

At 3 meters away from the camera the detection was still relatively well as the white space in the middle which represents the human could still be distinguished with relative ease. However some environmental noise did start to appear on the frame as seen on figure 5.13.

The amount of environmental noise present on the frame with the target 4 meters away from the camera completely overwhelms the tracking algorithm and as a result it didn't perform very well. This can clearly be seen on figure 5.14.

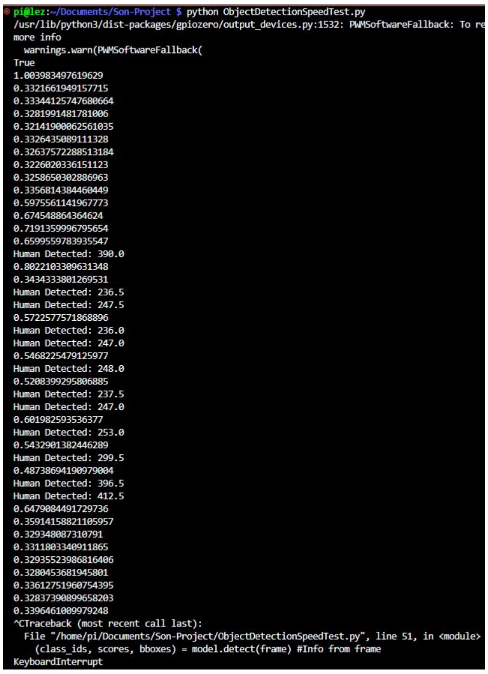
These results show that the background subtraction model is severely effected by environmental noise if the target is more than 3 meters away with little to noise present when the target is closer than 3 meters. This seems to suggest the algorithm struggles to lock onto targets further than 3 meters away and as a result cant distinguish between noise and the actual target in order to properly block out the noise. This means in order to achieve optimal tracking using the thermal camera the person must be within 3 meters of the device.

5.3.5 Tracking Speed of fusion tracking algorithm

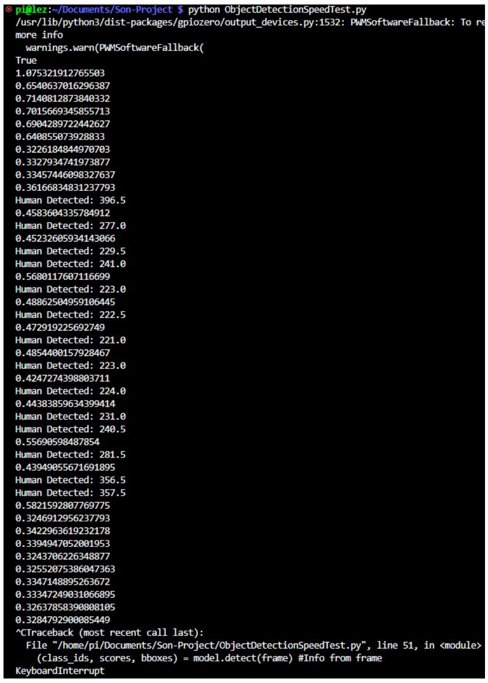
Testing Procedure

The tracking speed of the fusion algorithm can be measured by measuring the detection speed of the Yolov4-tiny and Background-subtraction object detectors. This will provide an indication of how fast an object can move across the screen while still allowing enough time for the algorithm to respond by turning the motor. To test this a timer will be placed in the algorithm to measure the time it takes for one loop to complete while an object is within the field of view of the cameras.

Experiment Results and Discussion



(a) Visual tracking algorithm speed test - 1

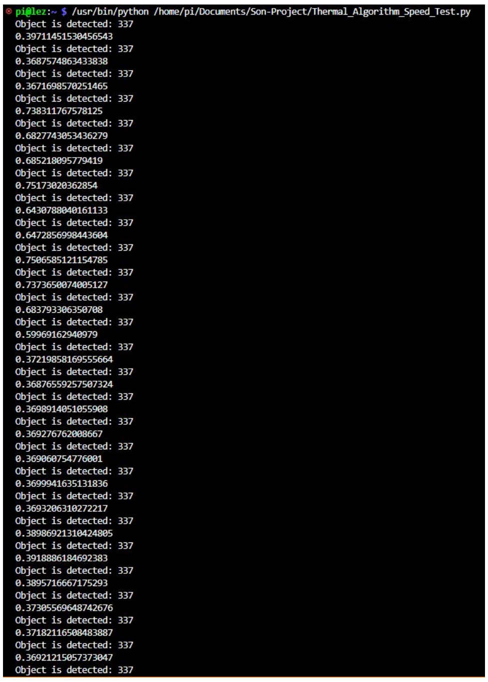


(b) Visual tracking algorithm speed test - 2

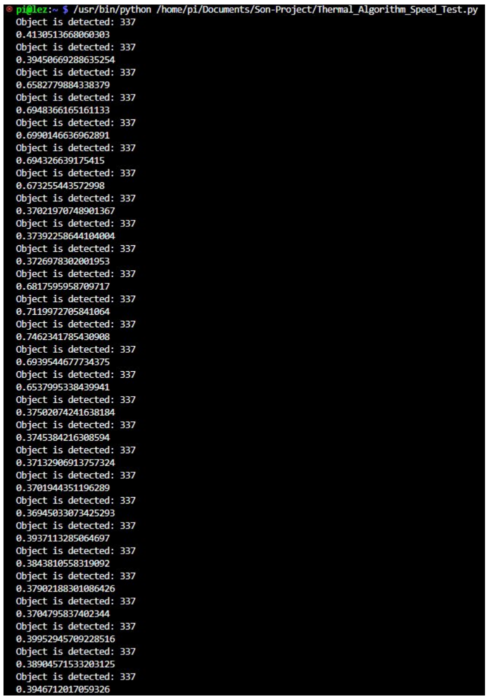
Figure 5.15: The two figures above show the time it takes for one loop to complete in the visual tracking algorithm while its detecting a person on the frame

From 5.15 we can see that when no person is detected on the visual frame it takes approximately 0.35s for one of the algorithm to complete. However when a person is detected on the visual frame the time it takes for one loop complete becomes approximately 0.5s. This result is consistent across both the test cases seen in 5.15.

This means that when a person detected on the visual frame the algorithm runs at a maximum of two frames per second which severely limits the algorithms ability to track objects moving at high speed. The poor performance of this algorithm is most likely due to the lack of a proper GPU on the Raspberry Pi 3B as CNN object detectors such as Yolov4-tiny are primarily designed to run on a GPU.



(a) Thermal tracking algorithm speed test - 1



(b) Thermal tracking algorithm speed test - 2

Figure 5.16: The two figures above show the time it takes for one loop to complete in the thermal tracking algorithm while its detecting a person on the frame

Unlike the visual tracking algorithm the background subtraction model used for tracking on the thermal camera runs at approximately 0.37s per loop consistently even when a object is detected. This equates to approximately 2.7 frames per second, which is also not ideal for tracking high speed objects. The algorithm does take longer to complete the initial couple of loops which is most likely due to the initial setup required at startup still effecting the run-time.

Overall both the visual and thermal tracking algorithms produced poor performance in terms of detection speed which means that the fusion algorithm which makes both is limited to tracking only slow moving targets.

5.4 Summary of the systems performance

The table below provides a summary of the systems performance in reference to the performance metrics laid out in section 5.2.

Table 5.1: Summary of the systems performance in reference to the performance metrics

Performance Metric	**System performance**
Environmental Noise Rejection	• Visual Camera - Excellent noise rejection with no false detections • Thermal Camera - Very susceptible to environmental noise
Detection range of the visual camera tracking system	• Minimum Range - 1m • Maximum Range - 5m
Performance of the visual tracking algorithm under varying light conditions	• Optimal lighting conditions - Excellent detection • Dim lighting conditions - Adequate/Good detection • Dark environment - No detection (Poor performance)
Detection range of the thermal camera object detection algorithm	• Minimum Range - N/A • Maximum Range - 3m
Tracking speed of fusion tracking algorithm	limited to tracking only really slow moving targets

Chapter 6

Discussion and Conlcusion

This chapter will discuss the results obtained in chapter 5 as well as make suitable concluding remarks on based based on overall success or failure of the project in reference to the main objectives set out in Chapter 1.

6.1 Discussion of Results

The main aim of the camera tracking system is to use data from a thermal- and visual-camera to improve the overall tracking of the system compared to using only a visual camera. The experiments performed in 5.3 were designed to test the extent of the improvement.

As seen on Table 5.1 the test done on the visual tracking system showed that it has excellent environmental noise disturbance and can perform excellent tracking under optimal and dim lighting conditions, but cant detect any human-objects in a dark room which is expected. The algorithm also has a detection range of between 1m - 5m, but is limited to tracking only slow moving targets due to the object detection running at maximum speed of approximately 2 frames per second.

Similar test were done on the thermal tracking system and they showed that the thermal camera is heavily effected by environmental noise when it has no clear target to lock onto. The result also showed that the thermal camera has a maximum detection range of approximately 3m before the target is to far away from the camera. Similar to the visual tracking system the thermal tracking system also produced a poor object detection speed of approximately 2.7 frames per second which means it is also limited to tracking slow moving targets.

As the Visual- and Thermal-camera fusion algorithm uses the visual and thermal tracking algorithm as part of its implementation means that it inherits all the properties and flaws of both algorithms. This means in optimal light conditions the fusion algorithm will perform similar to the visual tracking algorithm and in darker environments it will perform like the thermal tracking algorithm.

6.2 Conclusion

The main goal of this project was to develop a camera tracking system that combines data from a visual camera and a low-cost thermal camera. This goal was then further refined into three main sub-objectives outlined in section 1.3.

The proposed solution met all three main sub-objectives as the system was able to track at least one human object on a visual frame and by incorporating a low-cost thermal camera into the design the system was able to identify and track at least one person in poor lighting conditions. The third sub-objective was met through the design and construction of the mechanical housing and mount which housed all the main hardware components used during the project. Thus the overall project was a success.

Even though all the main objectives for this project were met the final design didn't perform up to a satisfactory level. The slow tracking performance of the system is not ideal as it means that the system cant be applied in any practical real-time scenario. The susceptibility of the low cost thermal camera to environmental noise along with its poor detection range suggest that it is not suitable for tracking applications and other options should rather be explored.

6.3 Recommendations for future work

This thesis implemented a basic camera tracking system using a visual- and low cost thermal-camera, however the final design left a lot to be desired for in terms of its performance.

Due to budget constraints the Raspberry Pi 3B was used in this project, however this severely impacted the speed of the Yolov4-tiny detection algorithm. To improve the performance in the future it would better to use a embedded device like the NVIDIA Jetson Nano which has a dedicated GPU as well as a faster CPU. This would increase the speed of the detection algorithms and allow the device to be used in practical real-time situations.

Another aspect of the design proposed in this thesis that negatively impacted the performance of the system is the Adafruit AMG8833 IR Thermal camera. One of the project objectives was to use a low cost thermal camera and as a result the Adafruit camera was used, however the 8x8 pixel resolution makes it very difficult to perform object detection on it. Going forward I would suggest using the Adafruit MLX90640 Thermal Camera which has a 24x32 pixel resolution. This module not only has a much better image resolution it also keeps the same compact design found on the Adafruit AMG8833 IR Thermal camera which makes it easy to integrate into the design.

If the design incorporates a more powerful embedded system like the Jetson Nano along with a better thermal camera like the Adafruit MLX90640 Thermal Camera it will allow for the use of a custom trained YoloV4-tiny object detection algorithm on thermal frames rather than using background subtraction methods. This would drastically improve the noise rejection and detection range of the thermal tracking algorithm. It would also be beneficial to include a lux sensor to better distinguish when to switch between and thermal tracking

The 3D Printed parts of the mechanical housing are very fragile and broke a few times during the implementation of this project. Thus for future work it would be best to construct the casing of a stronger material or re-design it completely in order to reinforce fragile parts.

Bibliography

[1] Lanny, "Block diagram of digital camera," Apr 2021. [Online]. Available: https://quotestrendingupdateau.blogspot.com/2021/04/block-diagram-of-digital-camera.html

[2] I. Morawsk, Y.-A. Chen, Y.-S. Lin, and W. H. Hsu, *NOD: Taking a Closer Look at Detection under Extreme Low-Light Conditions with Night Object Detection Dataset*, 2021.

[3] V. Lendave, "Guide to simple object detection using inceptionresnetv2," Jul 2021. [Online]. Available: https://analyticsindiamag.com/guide-to-simple-object-detection-using-inceptionresnet_v2/

[4] C. Kim, M. Oghaz, J. Fajtl, V. Argyriou, and P. Remagnino, "A comparison of embedded deep learning methods for person detection," *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2019.

[5] S. Verma, A. Soni, V. Mishra, V. Gupta, R. Krishnamurthi, and A. Kumar, "Smart automated system for classification of emergency heavy vehicles and traffic light controlling," *Autonomous and Connected Heavy Vehicle Technology*, p. 245–262, 2022.

[6] A. Berg, "Detection and tracking in thermal infrared imagery," Ph.D. dissertation, Linköping University, 2016.

[7] M. Hayk, "What is the thermal infrared range? and what are nir, swir, mwir and lwir?" 2021. [Online]. Available: https://www.ametek-land.com/pressreleases/blog/2021/june/thermalinfraredrangeblog

[8] D. Galar and U. Kumar, *Sensors and Data Acquisition.* Academic Press, 2017.

[9] B. Abidi, J. Heo, S. G. Kong, and M. Abidi, "Fusion of visual and thermal signatures with eyeglass removal for robust face recognition," 2004. [Online].

Available: https://www.researchgate.net/publication/4119610_Fusion_of_Visual_and_Thermal_Signatures_with_Eyeglass_Removal_for_Robust_Face_Recognition

[10] R. Pi, "Buy a raspberry pi 3 model b." [Online]. Available: https://www.raspberrypi.com/products/raspberry-pi-3-model-b/

[11] C. S. Africa, "Hkd raspberry pi camera mod 5mp." [Online]. Available: https://www.communica.co.za/products/hkd-raspberry-pi-camera-mod-5mp?variant=31916110217289

[12] "Adafruit amg8833 ir thermal camera." [Online]. Available: https://www.robotics.org.za/AF3538?search=thermal

[13] J. Hrisko, "Thermal camera analysis with raspberry pi (amg8833)," Jul 2021. [Online]. Available: https://makersportal.com/blog/thermal-camera-analysis-with-raspberry-pi-amg8833

[14] R. Yamashita, M. Nishio, R. K. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, p. 611–629, 2018.

[15] S. Balaji, "Binary image classifier cnn using tensorflow," Aug 2020. [Online]. Available: https://medium.com/techiepedia/binary-image-classifier-cnn-using-tensorflow-a3f5d6746697

[16] H. Gernsheim, *A concise history of photography.* Dover, 1986.

[17] S. T. McHugh, *Understanding photography.* Cambridge in Colour, 2017.

[18] B. Dipert, "Overcome these 6 problems with object detection," Feb 2022. [Online]. Available: https://www.edge-ai-vision.com/2022/02/overcome-these-6-problems-with-object-detection/

[19] R. Verschae and J. Ruiz-del Solar, "Object detection: Current and future directions," Jan 1AD. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frobt.2015.00029/full#:~:text=Object%20detection%20has%20been%20used,photo%20management)%2C%20and%20(vi

[20] N. Yadav, "Comparative study of object detection algorithms," *International Research Journal of Engineering and Technology (IRJET)*, Nov 2017.

[21] P. Soviany and R. T. Ionescu, "Optimizing the trade-off between single-stage and two-stage deep object detectors using image difficulty prediction," *2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, 2018.

[22] A. AbdulMalik, A. Khalil, and H. Ullah Khan, "Object detection and tracking using background subtraction and connected component labeling," *International Journal of Computer Applications*, vol. 75, no. 13, p. 1–5, 2013.

[23] H. Bergenroth, "Use of thermal imagery for robust moving object detection," Jun 2021. [Online]. Available: https://liu.diva-portal.org/smash/get/diva2:1578091/FULLTEXT01.pdf

[24] "Discovery-of-infrared-light-nasa." [Online]. Available: https://spaceplace.nasa.gov/review/posters/herschel/Herschel-ir-activity.pdf

[25] AGMGlobalVision, "When was thermal imaging invented and who." [Online]. Available: https://www.agmglobalvision.com/when_was_thermal_imaging_invented_and_who

[26] "Physical basics." [Online]. Available: https://www.infratec.eu/thermography/service-support/glossary/theory/#:~:text=The%20principle%20of%20infrared%20thermography,composition%20of%20its%20emitted%20radiation.

[27] "Your guide to industrial thermal imaging cameras." [Online]. Available: https://tameson.com/thermal-imaging-camera.html#:~:text=Thermal%20cameras%20detect%20infrared%20radiation,detected%20by%20the%20camera%20sensor.

[28] B. B. Lahiri, S. Bagavathiappan, T. Jayakumar, and J. Philip, "Medical applications of infrared thermography: A review," Jul 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7110787/#:~:text=IRT%20has%20been%20successfully%20used,fever%20screening%20and%20brain%20imaging.

[29] O. Engineering, "Top applications of thermal imaging cameras," Dec 2021. [Online]. Available: https://www.omega.com/en-us/resources/thermal-imagers-applications

[30] F. Morgado, "Thermal imaging in ai," Apr 2020. [Online]. Available: https://nilg.ai/202004/thermal-imaging-in-ai/

[31] R. Prajapati, A. Bosamiya, S. Dhoranwala, K. Bhatt, and P. Luhana, "Sensor fusion: A review of methods," Apr 2020. [Online]. Available: https://www.ijirmf.com/wp-content/uploads/IJIRMF202004010.pdf

[32] C. Tangemann, "Sensor fusion: Technical challenges for level 4-5 self-driving vehicles," Oct 2019. [Online]. Available: https://www.automotive-iq.com/autonomous-drive/articles/sensor-fusion-technical-challenges-for-level-4-5-self-driving-vehicles#:~:text=One%20significant%20issue%20is%20the,and%20in%20spatio%2Dtemporal%20alignment.

[33] "Stefan-boltzmann law." [Online]. Available: https://www.britannica.com/science/Stefan-Boltzmann-law

[34] "Planck's radiation law." [Online]. Available: https://www.britannica.com/science/Plancks-radiation-law

Appendix A

Additional Files and Schematics

A.1 Convolutional Neural Network Layers

Convolution Layer

The convolution layer is used to be perform feature extraction by making use of a special math operation called convolution and a non-linear activation function. The input to the convolution layer is a multidimensional array called a tensor. Within the convolution layer the tensor is convolved with a two-dimensional array called a kernel to produce a feature map.

Through the process of training various kernels are used in an attempt to find the one that works best for the given dataset. However the kernel size, number of kernels, stride and the amount of padding needed are hyper-parameters [14].

Pooling Layer

The pooling layer follows directly after the convolution layer and takes feature maps as an input. The pooling layer is a downsampling operation which aims to reduce the number of learnable parameters and subsequently help reduce overfitting of the data as well as reduce the hardware resources needed to run the algorithm.
Max pooling is the most common pooling operation used for CNN algorithms and works by looking at patches of input feature maps and producing the maximum value in each patch as its output. Another common pooling operation is average pooling which produces the average of the values in a given patch as an output. Average pooling has two main advantages, it reduces the amount of learnable parameters in the algorithm whilst also allowing the CNN to accept inputs of varying size [14][6].

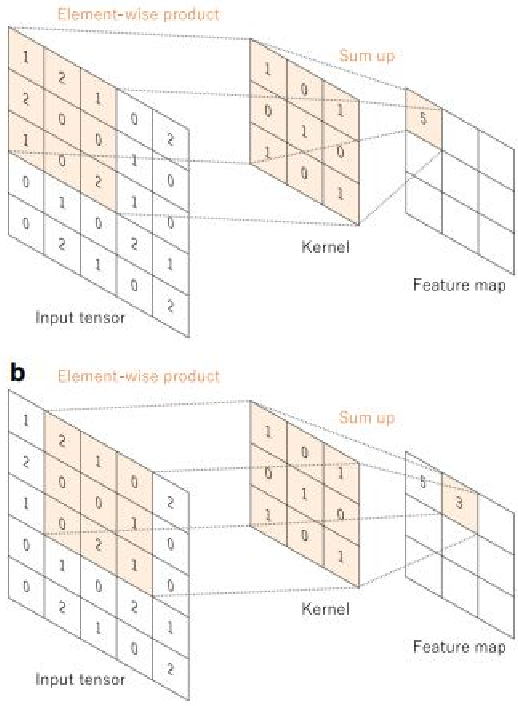


Figure A.1: A Diagram diagram displaying the process of generating a feature map by convolving an input tensor with a kernel [14]

Fully Connected Layer

The fully connected layer is the final layer in the CNN architecture and takes the output of the pooling layer as an input. The output data from the pooling layer is generally fed to the final output through a subset of fully connected layers. The last fully connected layer normally has the same number of outputs as the amount of classes the CNN is trained for whilst also having a different type of activation function. Typical activation functions used for the last fully connected layer include the ReLU, sigmoid and tanh functions [14].

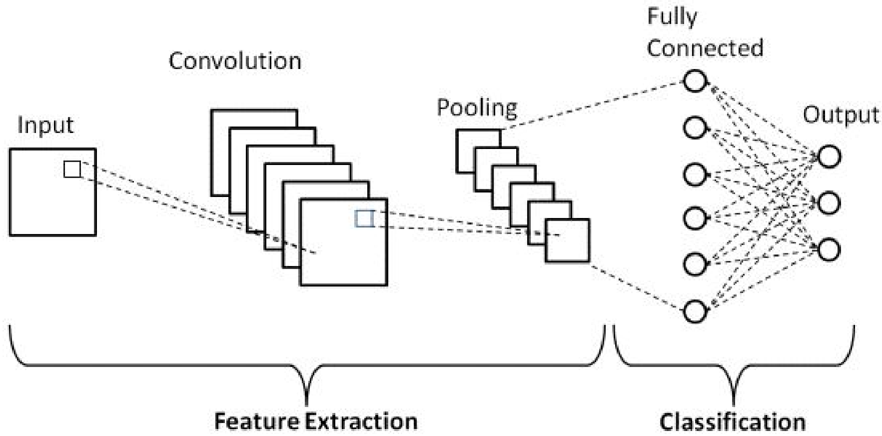


Figure A.2: Basic structure of the CNN Architecture [15]

A.2 Detailed Thermal Image Theory

According to the Stefan-Boltzmann law all matter in the known universe emits thermal radiation which is proportional to its absolute temperature. This relationship can also seen in the following equation.

$$E_R = \sigma T^4 \tag{A.1}$$

where E_R is the radiation energy emitted, σ is the constant of proportionality and T is the absolute temperature of the object [33]. Thus by utilising Planck's energy equation A.2 the wavelength of the radiated signal can be determined [34].

$$E = \frac{hc}{\lambda} \tag{A.2}$$

In the 1900s Max Planck also developed Planck's radiation law which proposes that an ideal object named a black body emits a spectrum of thermal radiation spanning a across the entire electromagnetic spectrum with hotter objects emitting more energy on the short-wavelength part of the spectrum (ultraviolet) and colder objects emitting more energy with lower wavelengths (infrared/microwaves) [34].

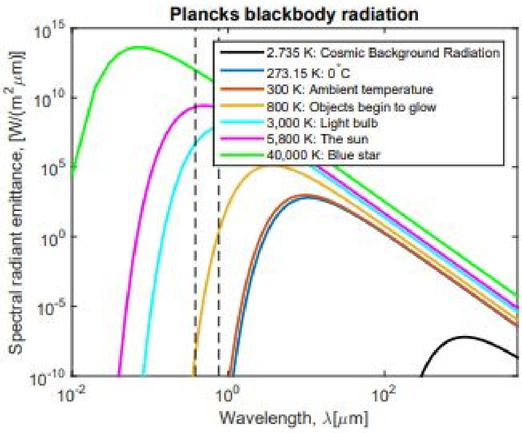


Figure A.3: A Diagram displaying the Spectral Density plots of radiation Energy emitted for black body objects at different temperatures (The dashed lines indicate the visual light spectrum) [6]

As the majority of objects on the planet including the internal temperature of humans and animals are in close to ambient temperature (300k/27°C) they will emit most of their thermal radiation in the infrared wavelength range as seen on fig: A.3, thus making infrared cameras extremely useful for object detection in low visibility situations.

Appendix B

Addenda

B.1 Ethics Forms

Application for Approval of Ethics in Research (EiR) Projects
Faculty of Engineering and the Built Environment, University of Cape Town

ETHICS APPLICATION FORM

Please Note:
Any person planning to undertake research in the Faculty of Engineering and the Built Environment (EBE) at the University of Cape Town is required to complete this form **before** collecting or analysing data. The objective of submitting this application *prior* to embarking on research is to ensure that the highest ethical standards in research, conducted under the auspices of the EBE Faculty, are met. Please ensure that you have read, and understood the **EBE Ethics in Research Handbook** (available from the UCT EBE, Research Ethics website) prior to completing this application form: http://www.ebe.uct.ac.za/ebe/research/ethics1

APPLICANT'S DETAILS		
Name of principal researcher, student or external applicant		Lezerick Owies
Department		Engineering Build and Environment
Preferred email address of applicant:		owslez001@myuct.ac.za
If Student	Your Degree: e.g., MSc, PhD, etc.	BScEng Mechatronics
	Credit Value of Research: e.g., 60/120/180/360 etc.	40
	Name of Supervisor (if supervised):	Jarryd Son
If this is a researchcontract, indicate the source of funding/sponsorship		N/A
Project Title		Enhanced Camera Target Tracking Using a Low-cost Thermal Camera.

I hereby undertake to carry out my research in such a way that:

- there is no apparent legal objection to the nature or the method of research; and
- the research will not compromise staff or students or the other responsibilities of the University;
- the stated objective will be achieved, and the findings will have a high degree of validity;
- limitations and alternative interpretations will be considered;
- the findings could be subject to peer review and publicly available; and
- I will comply with the conventions of copyright and avoid any practice that would constitute plagiarism.

APPLICATION BY	Full name	Signature	Date
Principal Researcher/ Student/External applicant	Lezerick Graeme Owies	L.G. Owies	18/08/2022
SUPPORTED BY	**Full name**	**Signature**	**Date**
Supervisor (where applicable)	Jarryd Son		18/08/2022

APPROVED BY	Full name	Signature	Date
HOD (or delegated nominee) Final authority for all applicants who have answered NO to all questions in Section 1; and for all Undergraduate research (Including Honours).			
Chair: Faculty EIR Committee For applicants other than undergraduate students who have answered YES to any of the questions in Section 1.			