Object Detection, Tracking, Re-identification, and Activity Recognition for Maritime Surveillance using Thermal Vision - STAGE II

Undergraduate graduation project report submitted in partial fulfillment
of the requirements for the
Degree of Bachelor of Science of Engineering
in
The Department of Electronic & Telecommunication Engineering
University of Moratuwa.

Supervisor:

Dr. Pehsala Jayasekara

Group Members:

A.M.N.D.Samarakoon(170537R)

W.S.D. Fernando (170168K)

K.P. Imbulgoda (170231A)

K.L.T. Ishara (170240B)

May 2022

Approval of the Department of Electronic & Telecommunication Engineering

....
Head, Department of Electronic &
Telecommunication Engineering

This is to certify that I/we have read this project and that in my/our opinion it is fully adequate, in scope and quality, as an Undergraduate Graduation Project.

Supervisor: Dr. Peshala Jayasekara

Signature: ....

Date: ....

DECLARATION

This declaration is made on May 29, 2022.

Declaration by Project Group

We declare that the dissertation entitled ”Road Sign, Traffic Light and Static Object Detection for Self-Driving - Stage II” and the work presented in it are our own. We confirm that:

- this work was done wholly or mainly in candidature for a B.Sc. Engineering degree at this university,
- where any part of this dissertation has previously been submitted for a degree or any other qualification at this university or any other institute, has been clearly stated,
- where we have consulted the published work of others, is always clearly attributed,
- where we have quoted from the work of others, the source is always given,
- with the exception of such quotations, this dissertation is entirely our own work,
- we have acknowledged all main sources of help,
- parts of this dissertation are yet to be published.

31/5/2022
Date

A.M.N.D. Samarakoon (170537R)

W.S.D. Fernando (170168K)

K.P. Imbulgoda (170231A)

K.L.T. Ishara (170240B)

DECLARATION BY SUPERVISOR

I/We have supervised and accepted this dissertation for the submission of the degree.

....	....
Dr. Peshala Jayasekara	Date

… # ABSTRACT

Object Detection, Tracking, Re-identification, and Activity Recognition-
Maritime Surveillance using Thermal Vision - STAGE II

Group Members: W.S.D.Fernando, K.P.Imbulgoda, K.L.T.Ishara, A.M.N.D. Samarakoon

Supervisor: Dr. Peshala Jayasekara

Keywords: Thermal Vision, Object Detection, Activity Detection, Tracking, Suspicious Activity, Object Re-Identification, Self-Stabilization, Wave Compensation

Maritime surveillance is a crucial element in the world, as most of the illegal activities are practiced in the maritime environment since it is a hotspot in world trade. More than 90% of the world trade is dependent on maritime transportation. If we go through history, maritime transportation is not a new thing for the world and it has been evolved throughout history and with the evolution, illegal activities related to the maritime domain also increased drastically. Due to this rapid growth, safety and security has become major issues

In the present world, many technological enhancements can be seen, and to ensure safety in the maritime industry, thermal IR camera systems with manual monitoring with the help of human intervention are present and used by related organizations like Navy and Coastal guards. Major concerns which affect maritime domain safety are terrorist activities, human smuggling, drug trafficking, and maritime pollution which cause much harm for thousands of million people throughout the world. Furthermore, we can identify some issues like illegal fishing which is related to the maritime environment which causes trouble for the fish industry as well as for nature. Therefore, monitoring the maritime environment at any time of the day in real-time is much more important and a proper and efficient method to fulfill that is required.

With this project, we propose a fully autonomous system using current technological advancements in robotics and machine vision to detect predefined objects, track them, identify actions through logical analysis, and to re-identify them(vessels) if we detect those objects (vessels) again. Also, for the camera, a self-stabilizing platform is proposed to address the complicated and unpredictable disturbances in the ship environment.

Many of the currently available systems are limited by some factors like the domain, speed, effectiveness etc. Most of the systems are operated in the RGB domain and it isa very common area of research. But there is a research gap in our area as we are dealing with Thermal IR images and implementing the system in the maritime environment.

The final solution of our project will run on inference hardware where we will feed the thermal video feed and it will predict the objects, localize, track, and re-identify if the object is a familiar image to the system. Furthermore, the stabilization platform will compensate the disturbances occurring in maritime environment.

Acknowledgement

We would like to express our special thanks of gratitude to our supervisors, Dr. Peshala Jayasekara and Dr.Ranga Rodrigo, for their endless guidance, support, and dedication towards the success of our project.

Further, we are thankful to our stage I members of this project for giving a tremendous support for completion of the stage II.

Dedication

To our families, friends, supervisors, and all others who supported us.

Abbreviations

RGB - Channels of a color image (Red, Green, Blue)

RNN - Recurrent Neural Network

AP - Average Precision

mAP - Mean Average Precision

SMD - Singapore Maritime Dataset

IR - Infrared

CMC - Cumulative match characteristic

I2C - Inter-Integrated Circuit

IMU - Inertial Measurement Unit

PWM - Pulse Width Modulation

SORT - Simple online and real-time tracking

TIR - Thermal IR

GAN - Generative Adversarial Network

IoU - Intersection over Union

Contents

List of Figures

List of Tables

1 INTRODUCTION

Protecting maritime resources, securing the border and avoiding more illegal scenarios like drug trafficking, human trafficking are the main border related security concerns in Sri Lanka in the maritime environment. These situations are increasing day by day and to overcome these situations we need to increase the security measures practiced around the border by updating the security protocols. The Sri Lanka Navy is the main peacekeeper between the borders of Sri Lanka and India and they operate maritime surveillance through day and night. Due to the difficulty to observe maritime environments at nighttime they have occupied resources like thermal infrared maritime surveillance systems under human intervention. But to perform the task in a more efficient manner there exists a need for an automated surveillance platform.
Most of the surveillance systems carries either a radar based or a vision based architecture. Though the radar technology provides accurate detection for longer range of distance, it comes with drawbacks like expensiveness and the performance get varied due to many factors. With the rise of machine vision systems and improved computer vision algorithms to perform tasks in high accuracy, research in building vision based applications for surveillance have become popular. Identifying maritime vessels, people and other objects, classifying the identified vessels as fishing boats, speed boats, ships etc. and tracking them are expected to be the key roles that a maritime surveillance system should be able to perform. In addition to them comprising the systems with features like re-identification and maritime activity detection helps to build a much superior maritime surveillance system. As RGB domain surveillance systems get effected by the lack of illumination at night and other low visibility weather conditions, our research works are carried out to build a maritime surveillance system in thermal vision.

1.1 Project Objectives

- Even Though maritime surveillance is a time consuming, fatigue and monotonous task, there are a very limited number of automated surveillance platforms out there specially systems with features like object detection and tracking, activity recognition, and vessel re-identification together.

- In order to develop object detection and tracking models, activity recognition models, vessel re-identification etc. quality datasets with features related to each model is essential. Though that's the case, currently there's no available dataset consisting of thermal maritime related images that can be used to train every above model.
- There are very few research works carried out for building activity recognition systems and vessel re-identification systems in thermal domain.

1.2 Project Scope

In order to build a system that will provide a solution for the above mentioned problem the deliverable have been identified as follows.

- Improved Object Detection & Tracking Model
- Model for Re-Identification for Vessels
- Activity Recognition System
- Improved in User-friendly interface
- Properly Annotated Dataset
- Camera Stabilization Platform
- Research Publication

1.3 Related Works

This section will provide a comprehensive survey on the main components of the whole system which includes stabilization platform, object detection, object tracking, vessel re-identification, action recognition.

1.3.1 Self Stabilization Platform

According to our research on this area, we found several platform types that was implemented before and which is related to our application. To build a complete system like this, several key researches has to be done like, research on sea conditions, research on the sea states and identifying our sea state, collecting IMU data around the Island where our system is deployed, boat behaviors, Environmental conditions likewise. Most of the researches we went through, they have collected datasets for their requirement and they analyzed the situation based on that. One such research is Design of Stewart Platform for Wave Compensation by Kristensen and Anders [31]. This was a large scale project to develop and build a Crane platform to stabilize or compensate against waves in the rough sea conditions. They here considered Stewart Platform as the most suitable parallel manipulator for their project which is a 6 dof platform having 6 prismatic actuators connecting two rigid bodies. They used this to handle heavy loads because of its parallel structure and control accuracy of the parallel manipulators are much high and they considered that factor as well. During the kinematics development for this, they considered a velocity and the force from the prismatic joints. We went through another research [6] and this was designed for a wind turbine mounted on a ship/vessel. Through the wave compensation they tried to protect their wind turbines from collisions due to rough sea conditions. For a typical wave compensation, we have to deal with the types of motions like roll, pitch, yaw, heave, sway, and surge [11] [10].

1.3.2 Object Detection

Object detection in our context is employed for iden- tification and spatial localization of object such as such as ships, boats, unidentified floating objects, humans, etc. With rise of deep learning techniques, many have incorporated them in building object detector achieving significant success.[34] Usage of the concept of the bounding box proposals have been a popular method to accomplish this task.[26] Then the method of using keyponis to localize objects was introduced[23], which later extended to the concept of detection the center-point of the object[8]. This architecture is known as CenterNet and has shown an improved inference speed due its one stage object detection nature. YoloV4[5] is also another object detector which shows a similar inference speed with improved accuracy. These architecture are supposed to perform in RGB domain.

Object detection in thermal domain was not a well explored reaserch area. Image processing techniques like C-means clustering[19], pixel intensity histogram segmentation etc. have been used for detection in thermal domain.[1]. But as the object detectors presented in RGB domain provide a much superior performance usage of such detectors in thermal domain by training them with a thermal dataset has also been explored.[20].

1.3.3 Object Tracking

Methods of continuously localizing a target using computer vision techniques have been researched to build a system to track vessels. Semi-supervised discriminative tracking approaches has been utilized for single object tracking scenarios.[13] As our system intended to have a real time input, it has to be capable to track multiple objects belonging to multiple categories. Therefore Multiple Object Tracking (MOT) algorithms should be employed for the system. Also by considering that the enhancement of detection quality directly effects the tracking performance faster R-CNN[40], Mask R-CNN[46], SSD[41] have been used for object detection steps while for the tracking step Hungarian algorithm[4], Multi Layer Perceptron[21] have been used before.

For object tracking in maritime context, Szpal *et al*. has developed a solution by employing background subtraction and real-time approximation of level-set-based curve evolution to distinguish the outlines of the vessels to track them[37]. In SiameFPN which is a deep learning based tracking pipeline they have used a combination of modified Siamese Network with multi FPNs to solve the above challenge[36]. To perform multiple vessel tracking Jie *et al*. has utilized improved YOLOv3 with Deep SORT algorithms[18].

1.3.4 Action Recognition

The area of activity detection has experienced greatly improved performance in recent years owing to the introduction of deep learning-based approaches. Though the core intention of activity recognition and detection is similar, in implementation wise and presentation wise they differ from each other. Activity recognition seeks to identify the actions taking place in an entire video, such as in [42] whereas activity detection seeks to localize that activity in time and/or space, such as in [29][12]. For our application as we are intended to localize and display it in a GUI we are much interested in developing an action detection method.

There are very limited number of action recognition methods are in thermal domain. Batchuluun *et al*. has presented a CNN-LSTM based method using joint and skeleton information[2]. As they have used a CycleGAN to perform image restoration and the removal of a halo effect to improve the extracted silhouettes of humans, the system is hard to use for real time action recognition.

1.3.5 Object Re-Identification

In person re-identification research area, methods like feature learning[27], distance metric learning[7], subsapce learning[33] have been utilized. But using person re-identification systems for our context would not be much feasible as viewpoint variation makes a huge impact on the accuracy of vessel re-identification. Due to the lack of research works for our domain we have broaden our focus area by going through some vehicle re-identification methods as well. In these domains research works, a separate mechanism for viewpoint estimation have been proposed along with methods for re-identification[45]. But performing vessel re-identification in thermal domain provides much harder challenge than any of the research works above. Without being relying only on thermal vision some have research on using both RGB and thermal domains to perform re-identification[29], due to the problem of having limited features in thermal imagery.

2 METHODOLOGY

We will discuss our project overall methodology in this section.

2.1 Project Overall Architecture

The below diagram will show the overall architecture of the project. The video feed will be taken from the FLIR M232 thermal camera. In order to take a clear and a stable thermal video feed the camera will be mounted on a wave compensated stabilization platform. Then the thermal video will be fed to the object detection model to detect vessels, maritime activities etc. The results of it will be provided to the tracking model, vessel re-identification model and activity detection model. Using the activity detection model, the detected actions will be classified as swimming, skiing etc. and the results will be input to the GUI. The detected vessels will be further identified using the vessel re- identification model and also the results output from that is fed to the GUI. Using the GUI and the additional details displayed on it, the customer will be able to perform the maritime surveillance in a more advanced and efficient manner.

2.2 Thermal Camera

The starting point of our system is the FLIR M232 Thermal IR camera. It has a small light weight design and specially built for maritime environments with features like IPX6 water resistance and salt and mist protection (IEC 60945). Therefore, this camera can be used in maritime environments safely and reliably.

The camera has an H264 IP Video stream that can be observed via a web-based user interface and can be accessed separately as a RTSP video stream. Firstly, this video feed is obtained in real-time from the FLIR M232 Thermal IR Camera. Then the obtained video feed is fed to the object detection and tracking model.

The FLIR M232 camera has a built-in pan tilt control system. This control system has a tilt range of +110° to -90° and a continuous pan range of 360°.This in-built control

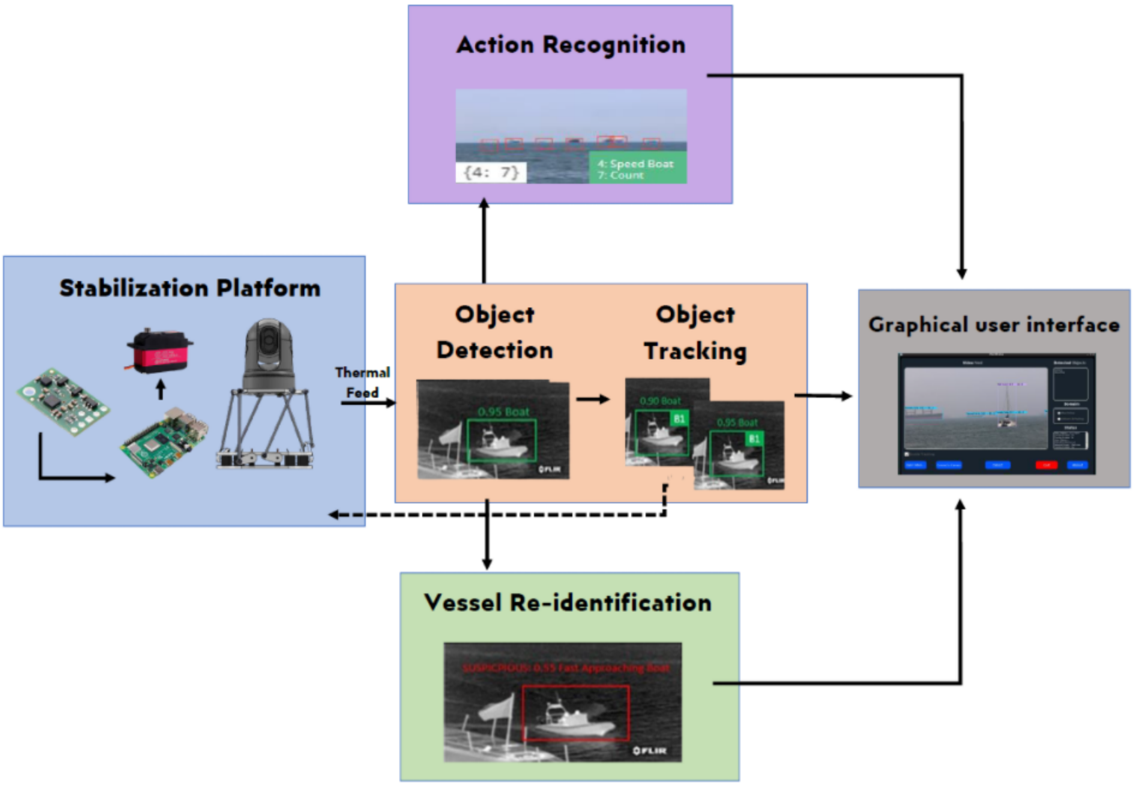


Figure 1: Architecture of the overall system

system can be used for tracking the object by physically moving the camera via pan and tilt. As only one object can be tracked by moving the camera physically, out system will track a specific user-defined object.

2.3 Self Stabilization Platform

The platform is the physical stabilizer for the camera video stream. The physical disturbance is high in the maritime environment and to address them, a proper wave compensation method that is highly non-linear is required. This is a complicated scenario as the motions are highly unexpected and non-linearity is much strong. In a maritime environment, a vessel is subjected to 6 types of disturbances like roll, pitch, yaw, sway, heave and surge.[6] To compensate for these motion types, we chose 6 degrees of freedom mechanism called Stewart Platform.

We developed two different modes for the Stewart Platform as given below,

- Manually Controllable mode.

- Self Stabilizing mode.

In **Manually Controllable mode**, the platform has the capability to change its orientation and position according to the user input orientation. This is a open loop system and flow and the results will be discussed more in the next sub sections. In **Self Stabilization mode**, the platform has capability to self stabilize its stage with respect to the world frame. This is a closed loop control system and here we used a PID controller to mitigate the error.

Top view block diagram for Manually controllable mode

In 1st block, orientation or position or combination of both is given to the system. Then it passes through the inverse kinematics block and calculate the joint parameters for the 6 motors in rad. These parameters are mapped to PWM signals to control the 6 servo motors. In the end, the platform stage changes its orientation and position much precisely according to the input provided.

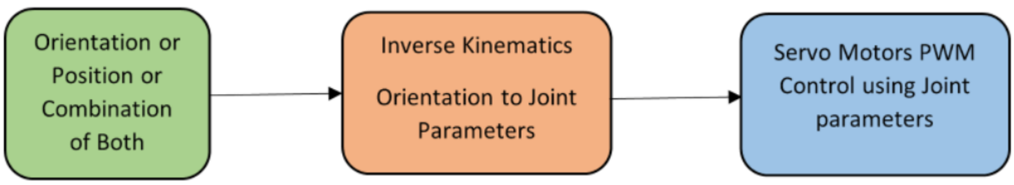


Figure 2: Manually Controllable Mode : Architecture

Several issues were faced by us during the development of above mode and we found solutions as given below.

Issues	**Solutions**
Latency of the system was high	Converted the python script which was initially programmed to C++ language and abled to increase the frequency from 5 Hz to 100 Hz.
Sudden continuous rotation of servo motors due to power excitation and loss of reference	Abled to solve it by using decoupling capacitors for the servo driver module.

Table 1: Issues and Solutions : Manually Controllable Mode

Top view block diagram for Stabilization Platform

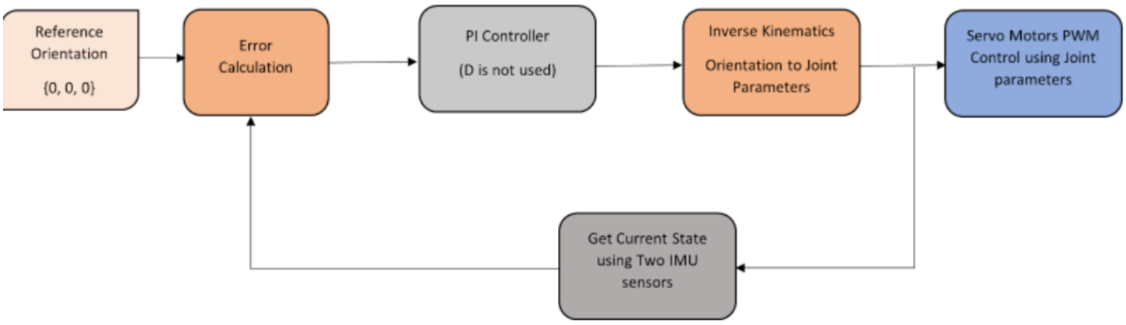


Figure 3: Stabilization Control System

Here we consider three frames which are **global frame, stage frame, base frame**. Here to measure the orientation with respect to these frames, we use two Inertial measurement unit sensors (IMU). One IMU sensor is mounted on the stage and the other one is mounted on the base. IMU which mounted on the stage provides the orientation of the stage with respect to the global frame. The other IMu provides the orientation of the base with respect to the global frame. Using these two orientation and position, we obtain the orientation of the stage with respect to the base. Here we used **Madgwick** filter for sensor fusion of data generating from the accelerometer and the gyroscope of the IMU sensor.

First, we calibrate the IMU sensor to the given local orientation and we consider the reference orientation roll, pitch and yaw angles as $\{0,0,0\}$ rad. We obtain the current state changes through the IMU sensors. When an external factor affects to the stability of the platform, the orientation of stage with respect to the global frame changes. We consider this as an error and through the closed loop control system, we correct that error with an frequency around 70 Hz in the practical scenario (theoretically it is 100 Hz). The main component which correct the error is our PI controller. We used trial and error moethod to determine suitable parameters for this model.

Issues	Solutions
High Latency Issue	Optimizing the program and calibrating the PI parameters.

Table 2: Issues and Solutions : Self Stabilization Mode

Inverse Kinematics

To address the inverse kinematics[9] problem of the Stewart manipulator, we calculated the rotary actuator joint parameters of the Stewart Platform. We will provide the whole procedure which we are following when calculating those.

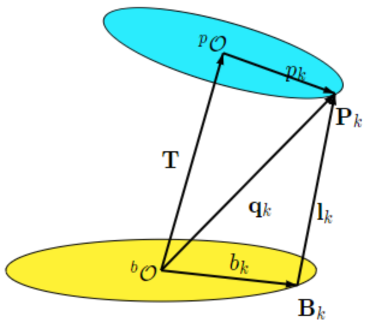


Figure 4: Sketch of the Stewart platform[9]

$$\textbf{Algorithm return} \longrightarrow \alpha_k,\ k = \{1,2,3,4,5,6\}$$

Base Anchor positions, stage anchor positions, Servo horn anchor positions are Bk, Pk, and Hk respectively. Here k = 1,2,3,4,5,6

Through the above vectors, we can find the rod length, horn length of each. The vector h is perpendicular to the servo shaft sk and rotated by an angle α_k. Refer to the Figure 5. When we look from above, we can observe β_k which is based on the placement of motors for the local x-axis.

Calculating anchor Hk which is a variable

Following equations are derived based on the Hk, Bk, and Pk coordinates.

Through the above equations, we can substitute the value of Hk which we previously found below,

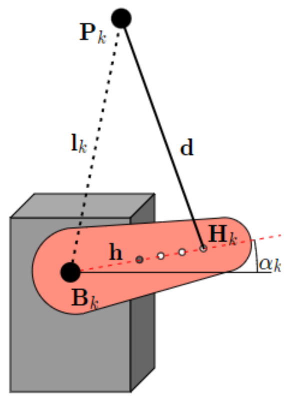


Figure 5: Servo Motor[9]

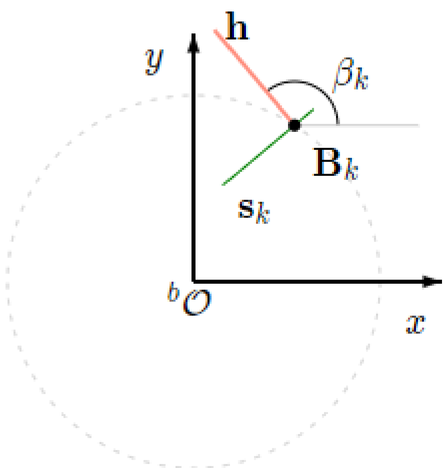


Figure 6: Top View[9]

$$
\begin{aligned}
\mathbf{H}_k &= \mathbf{B}_k + \mathbf{R}_z(\beta_k)\mathbf{R}_y(-\alpha_k)\begin{pmatrix} |\mathbf{h}| \\ 0 \\ 0 \end{pmatrix} \\
&= \underbrace{\mathbf{B}_k + |\mathbf{h}|\begin{pmatrix} \cos(\alpha_k)cos(\beta_k) \\ \cos(\alpha_k)sin(\beta_k) \\ \sin(\alpha_k) \end{pmatrix}}_{=\mathbf{h}}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{H}_k &= \mathbf{B}_k + \mathbf{R}_z(\beta_k)\mathbf{R}_y(\pi-\alpha_k)\begin{pmatrix} -|\mathbf{h}| \\ 0 \\ 0 \end{pmatrix} \\
&= \mathbf{B}_k + |\mathbf{h}|\begin{pmatrix} \cos(\alpha_k)cos(\beta_k) \\ \cos(\alpha_k)sin(\beta_k) \\ \sin(\alpha_k) \end{pmatrix}
\end{aligned}
$$

Now using Trigonometric Identity

$$\begin{aligned} |\mathbf{h}|^2 &= (\mathbf{H}_k - \mathbf{B}_k)^T(\mathbf{H}_k - \mathbf{B}_k) \\ |\mathbf{d}|^2 &= (\mathbf{P}_k - \mathbf{H}_k)^T(\mathbf{P}_k - \mathbf{H}_k) \\ |\mathbf{l}_k|^2 &= (\mathbf{P}_k - \mathbf{B}_k)^T(\mathbf{P}_k - \mathbf{B}_k) \end{aligned}$$

$$\begin{aligned} |\mathbf{l}_k|^2 - (|\mathbf{d}|^2 - |\mathbf{h}|^2) &= 2\mathbf{B}_k^T\mathbf{B}_k - 2\mathbf{B}_k^T\mathbf{H}_k - 2\mathbf{B}_k^T\mathbf{P}_k + 2\mathbf{H}_k^T\mathbf{P}_k \\ &= 2(\mathbf{H}_k - \mathbf{B}_k)^T(\mathbf{P}_k - \mathbf{B}_k) \\ &= 2|\mathbf{h}| \begin{pmatrix} \cos(\alpha_k) cos(\beta_k) \\ \cos(\alpha_k)\sin(\beta_k) \\ \sin(\alpha_k) \end{pmatrix}^T \mathbf{l}_k \\ &= 2|\mathbf{h}|\sin(\alpha_k)\mathbf{l}_k^{(z)} + 2|\mathbf{h}|\cos(\alpha_k)(\cos(\beta_k)\mathbf{l}_k^{(x)} + \sin(\beta_k)\mathbf{l}_k^{(y)}) \end{aligned}$$

$$e \cdot \sin\varphi + f \cdot \cos\varphi = \sqrt{e^2 + f^2}\sin(\varphi + \operatorname{atan2}(f, e)),$$

We can derive the following equations

$$\begin{aligned} e_k &= 2|\mathbf{h}|\mathbf{l}_k^{(z)} \\ f_k &= 2|\mathbf{h}|\left(\cos(\beta_k)\mathbf{l}_k^{(x)} + \sin(\beta_k)\mathbf{l}_k^{(y)}\right) \\ g_k &= |\mathbf{l}_k|^2 - (|\mathbf{d}|^2 - |\mathbf{h}|^2) \end{aligned}$$

Finally, the IK problem is reduced to the following,

$$\begin{aligned} g_k &= e_k \cdot \sin\alpha_k + f_k \cdot \cos\alpha_k \\ &= \sqrt{e_k^2 + f_k^2}\sin(\alpha_k + \operatorname{atan2}(f_k, e_k)) \\ \Leftrightarrow \quad \frac{g_k}{\sqrt{e_k^2 + f_k^2}} &= \sin(\alpha_k + \operatorname{atan2}(f_k, e_k)) \\ \Leftrightarrow \quad \alpha_k &= \sin^{-1}\left(\frac{g_k}{\sqrt{e_k^2 + f_k^2}}\right) - \operatorname{atan2}(f_k, e_k) \end{aligned}$$

Using the above motor angle vector, we can control the motors and through that we can achieve the desired orientation and position.

Inertial Measurement Unit Sensor : MPU 9250

- To calculate the state of the Stewart Platform, we require a measurement. Here we consider the IMU readings as the state of the Stewart Platform for our control system.

- Here we read the raw data which is Gyroscope data, Accelerometer data, and Magnetometer data from the IMU sensor, we first calibrate it to our requirements and pass that data through an IMU filter to estimate the Quaternion Representation and reduce the noise.

Arduino Mega

- The main processing unit for our platform is the Arduino Mega which has a ATmega2560 chip. Here we use this to calculate IMU estimates, IK problems, and to run the open loop and closed loop control systems.

Servo Motors

- Motors which we have used are servo motors and we use a servo controller to function those motors through PWM.
- We used I^2C to establish communication between Motors and the Servo Motor Driver.

Servo Driver : PCA 9685

- This was used for precise control of the servo motors through PWM signals.
- Through this module, we can control total 16 servo motors through the I2C interface.
- Servo Driver was handled by the Arduino Mega which we used as central control unit.

2.3.1 Software Stabilization

With the idea of gaining more stabilization on the camera video feed, software stabilization methods also tested. We tested on several software stabilization methods and to select a method,several evaluation metrics are used after referring several research works. Following are the metrics used for the evaluation.

1. Interframe Transformation Fidelity (ITF)
2. Interframe Similarity Index (ISI)
3. Average Speed (AvSpeed)
4. Average Average Percentage of Conserved Pixels (AvPCP)

Using these metrices, we evaluated 2 software stabilization methods, called point feature mapping and mesh flow stabilization.
The motion of video feeds may differ from each other with the environment. Therefore, it is difficult to keep an exact threshold value for each of these metrics, though we can calculate these metric values for each video. To solve this, a stable video feed was taken without creating any motion on the camera. It is the optimum video one can taken for the given environment.Then intentional motion was created on the camera in the same environment and obtained a video feed.Then that shaked video was sent into our software stabilization methods and calculated the metric values.Since then the optimum reference and shaked video metric values are available, we calculated the percentage of stability achieved by that software stabilization method.
There we have calculated how much amount of stability can be achieved by that software stabilization method as a percentage.

2.4 Object Detection

The thermal IR video feed obtained from the thermal camera then must be sent through an object detection stage to identify the instances of various maritime objects such as ships, boats, buoys, etc. As per the requirements of the system, the detection task should be carried out in real-time. The object detection section of the system should output the instance types and information required to set bounding boxes enclosing the object instances present in the image. The output of the detection stage will be then used in action object tracking, re-identification, and action detection stages of the system. Due to the high accuracy and inference speed of deep learning-based object detection is currently the most popular choice for this task.

This section of the project was completed by the stage one team of the project. After evaluating multiple deep-learning-based object detectors, the stage one team has chosen CenterNet[44] as the best available choice at the time according to selection metrics accuracy and inference speed.

	SSD	$YOLO-V3$	$CenterNet$
$Backbone$	VGG-16	DarkNet-19	DLA-34
$Stage$	Single	Single	Single
FPS	19	45	61
$f-mAP@IoU=0.5$	28.8	21.6	41.6

Table 3: Alternative Object Detection Algorithms

- In the stage 1 process, they have finalized their object detection models as Centernet[44]
- In our pipeline, they used DLA-34 as the model backbone. Centernet is very useful when operating in real-time as it is fast, and the computational cost is less when considering the memory likewise. It classifies objects and in regression, it detects the bounding box as a point. That is the main reason for its less computational complexities.
- In the feasibility study, we found some alternatives like YoLo V4[5] for the object detector. According to the results we obtained after evaluation, we finalized the prevailing system as our system as the comparison results were not significantly different.

2.4.1 Thermal Martime Object Detection

Thermal maritime object detection is a key deliverable in our project. Due to the unavailability of a dataset in thermal maritime domain, the stage one has evaluated this parts separately.(i.e,thermal object detection and maritime object detection as separate parts)
But since in our stage we created our own dataset in thermal-maritime domain. Therefore we could evaluate the given deliverable as it is. The obtained results are presented under results-object detection section section.

2.5 Object Tracking

Object tracking is a fundamental task in the computer vision domain. An object tracker should be capable of localizing a specific object throughout a provided video stream by maintaining a sufficient mAP score and an FPS value. In our system, the output of the object detection stage is fed into the object tracking stage. The object tracking section of the system then should localize the specific maritime object throughout the video stream as mentioned before. This should be carried out in real time. This introduces some constraints and requires making a trade-off between accuracy and the minimum significant delay that affects the FPS value.

This section of the project also has been completed by the stage one team. After evaluating multiple tracking methodologies they have selected the SORT tracking algorithm to provide the best performance in the required task.

As stage II team, we evaluated several models as well like Siamese tracking [3] and Deep Sort [39], but the main problem in those models were that, they cannot be operated at real time as we require. The reason is, these models are trainable models and the processing speed of them is low. When we consider SORT [4], it is an algorithm and it can be operated at real-time.

2.6 Action Detection

This section is describing the pathway to selecting a methodology and the selected methodology for action detecting task in the maritime environment.

2.6.1 Pathway to Selecting a Methodology

Finding a methodology for this task was highly challenging due to the constraints like camera image/ video quality and the availability of datasets with required actions. The state of art action detection algorithms expects high quality images to recognize the actions because action detecting requires detecting poses of the objects. But the cam-

era, FLIR M232 only is providing the resolution 320 x 240, which is not sufficient to proceed with the state of art action detection methodologies because valuable visual information is compromised by the low-resolution images. The next challenge was there were no publicly available dataset for particular actions in maritime environment.
After considering the above constraints, we selected a logical algorithmic based methodology to recognize maritime actions.

2.7 Methodology Architecture

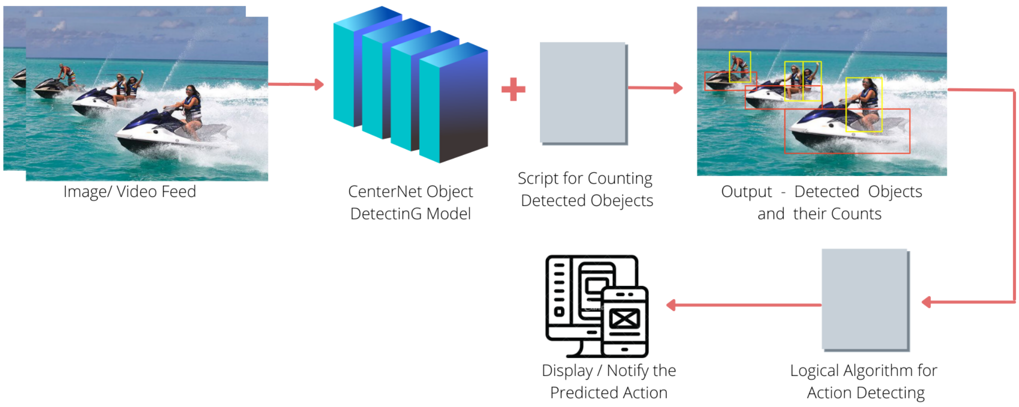


Figure 7: Action Detection Methodology Architecture

As displayed in the above architecture, the image or video feed provided by the camera will send in to the CenterNet object detection model and then the script that we have written output the count of the detected objects in each class. Then the output from the CenterNet and the script is sent into the logical algorithm that we have developed based on our action requirements. Finally, the algorithm predicted output of the action is displayed in the user interface which we have developed.

2.8 Progressing on Methodology

To detect maritime actions, after considering the constraints, requirements, and the resources we recognized several actions:

- A person is requesting help

- A Person is swimming
- A Person is jet skiing
- A Person is surfing.
- A buoy is floating

In addition to that we are detecting the suspicious action(when there is an unusual human count in a vessel/ship) and alert about them to the vessel security.

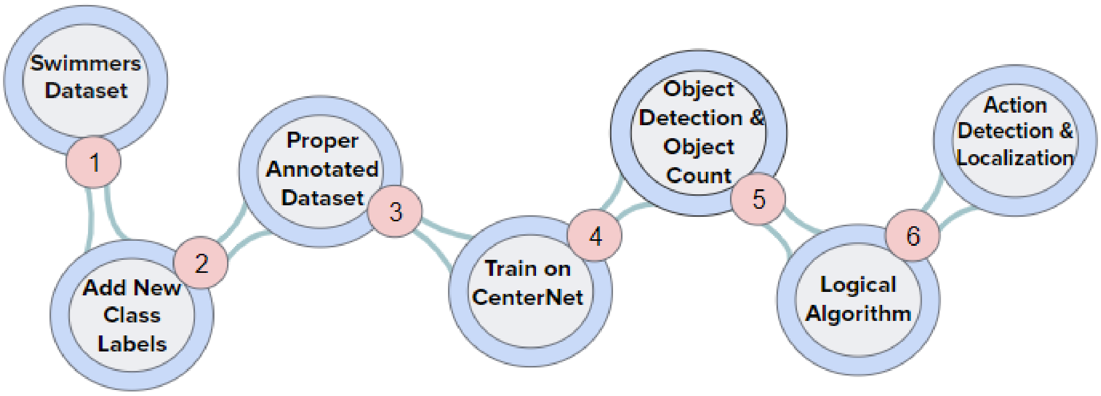


Figure 8: Action Detection Problem Solving Path

Above figure showing the approach that we followed on. We started to proceed with a dataset called swimmers publised on 2021 to detect maritime actions.It had only a single class named as swimmer. But after analysing the dataset, we added new labels to the dataset which are important on our task and finished the annotations.

The added new labels to the swimmer's dataset are;

- Help Requesting Person
- Safety Jacket
- Person in water
- Person not in water
- Surfboard
- Kayak
- Boat

- Sailboat
- Speed Boat.

After pre-processing the recreated swimmers dataset, it was trained on the CenterNet object detector and then obtained the predicted the bounding boxes and object count in video frames. Then they were sent into a logical algorithm. This logical algorithm is using Intersection over union value of the detected object's boundinng boxes to recognize and localize the actions. The output of the logical algorithm is capable of detecting and localization the action.
The main challenge in action detection was there were no any publicaly available dataset for martime actions. As a solution for that we tried to merge differnt datasets with different objects and then we used logical algorithm to predict.
Though our full system pipeline sholuld be on thermal domain, due to the unavailability of a dataset in thermal domain, we could not able to do the action detection on thermal domain for all the selected actions. But after our data collection and annotation we trained our thermal dataset for the same labels of the swimmers dataset and it peprformed well on the available actions(suspicious human count) in our dataset. Therefore, this methodology can directly apply for thermal domain when there is a proper dataset with required lables.

2.9 Vessel Re-identification

The idea of implementing the re-identification model is to identify the vessels that are common to a certain environment and identify intruders entering there, which is essential for maritime surveillance.Due to the unavailability of dataset containing thermal images of vessels which qualified to train a re-identification model and lesser amount of research work, the domain was decided to be changed to identify vehicles.

2.9.1 Model selection

The options considered for the vessel re-identification are - Strong Baseline for Vehicle reidentification - Marine vessel re-identification: a large-scale dataset and global-and-local fusion-based discriminative feature learning

The major pros and cons identified regarding the re-id models are mentioned in the table below

Strong Baseline for Vehicle reidentification[16]]	**Marine vessel re-identification: a large-scale dataset and global-and-local fusion-based discriminative feature learning [33]**
Verified state of the art performance on vehicle re-identification	Dataset and source code not publicly available for verification
Provides good results when trained even on synthetic data	Lack of data in thermal domain

After analysis of the options,among the research work done related to this area, The Strong Baseline Model[16] for Re-Identification was identified as the most suitable state of the art method for our application.

The first reason to consider this model is that the authors have identified the least amount of proper dataset available for the domain and have provided a solution to overcome that. In order to improve the accuracy of the model trained with small actual data, the authors have utilized an augmented synthetic dataset which is triple the size of the actual dataset. Then to reduce the error occurring by the domain difference, they have introduced a domain generalization method known as Mixstyle which adjusts the mean and the standard deviation of the actual and synthetic dataset. The model has occupied the RestNet model as the backbone, to handle the feature extraction method. Multihead attention mechanism has been utilized to identify single features extracted, which will also handle the viewpoint estimation. Also the priority for each identified feature will be decided by the above mechanism. Identification and metric losses were used during the training process of the model. Cross entropy loss with label smoothing was selected as the ID loss which tends to reduce intra class variance and to increase inter class variance. As the metric loss the authors have used supercon loss which is some kind of an extended version of triplet loss trained over the multiview batch consisting of real and randomly augmented images.

2.9.2 Modifications

The model was not designed by the authors for real time re-identification. It was instead designed for inferring the re-identification ids on images stored on a local disk. Initial startup time and inference time combined took approximately 3 minutes on a system with Nvidia GTX 1650Ti Max-Q GPU.The overall data flow was modified and initial galley image computation results were stored in a file to be loaded loaded at start up. These modification decoupled start up time and inference time and the load time and a image set inference time was brought down to 4 seconds and 250 milliseconds respectively for the same system mentioned before.

2.10 Data-sets

Datasets are an integral part of any machine vision application. They are used to train, test and evaluate models used for machine vision applications.With the advancement and increase of the popularity in machine vision domain, number of publicly available datasets has been increased considerably. Even so,over the course of the projects there were many instances our requirements were not met by the available public datasets. There is a general unavailabilty in maritime thermal domain datasets and it was a motivation in making 'Developing a thermal maritime dataset' a deliverable in our project. Many publicly available dataset were used by us and when there were shortcomings in the aavibale datasets we utilized alternate means or created or modified existing datasets.

2.10.1 Utilized publicly available datasets

Several public datasets in different domains were utilized for training and testing models during the course of our project. Those are

- Veri 776[27]
- RGBN300[15]
- RGBNT100[15]

- Swimmers dataset [30]
- FLIR self driving dataset
- Singapore Maritime Dataset
- Thermal Image Super-resolution dataset[35]

2.10.2 Created and modified datasets

There are several datasets developed during this project.

The 2 datasets created by us were captured with the FLIR M232 camera that was utilized throughout the project. After evaluating the options available Computer Vision Annotation Tool (CVAT) was chosen as the tool to be used for annotation due the availability of a web platform and many supported annotation features.

Maritime Thermal IR (MTIR) dataset

Dataset consists of thermal images covering the maritime environment which is one of the main deliverables of the project. It is suitable for the tasks of object detection and tracking, maritime activity detection. Contains 3978 annotated thermal IR images covering 18 different sceneries in maritime environment.

Figure 9: Sample images from MTIR dataset

Vehicle Thermal IR (VTIR) Dataset

As maritime re identification data are not currently available and could not be developed with the available resources, to train the re identification model with respect to our project's camera video feed and to demonstrate the re-identification functionality of the the complete system in re-identification , we created Vehicle Thermal IR (VTIR) data set with the FLIR M232 thermal camera. The dataset contains 2918 images, covering more than 5 viewpoints of 5 vehicles in thermal IR domain.

Figure 10: Sample images from VTIR dataset

Modified Swimmers dataset

Dataset covering maritime activities like surfing, swimming, skiing etc. for the model of activity detection. The images are already taken from the swimmers dataset, but we had to manually annotate the labels according to our requirement. Contains 4790 images in RGB domain.

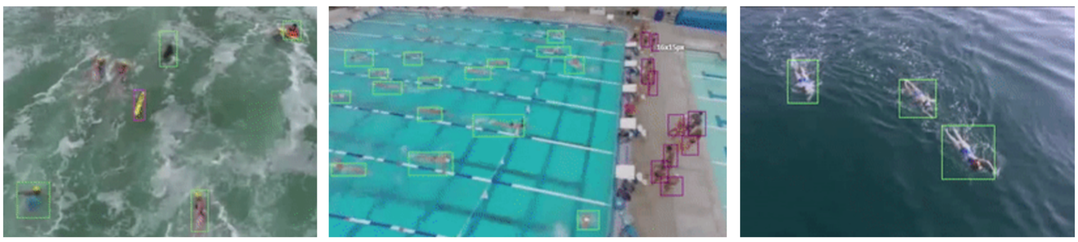


Figure 11: Sample images from Modified swimmers dataset

2.11 User Interface

As the last step of our project, we have supposed to develop a user-friendly interface to display the detected objects, actions and the alert about the vessel re-identification. Currently we have chosen PyQt5, the python binding of Qt which is heavily used in software development and runs on platforms supported by Qt which includes Windows, Linux, macOS, and Android. PyQt is licensed under the GPL v3 and the Riverbank

Commercial license. The reasons behind the selection of PyQt over many other frameworks can be highlighted as below:

- Offline framework as it is robust to network delays
- Easy to handle after installing locally
- Can be directly integrated with the algorithms developed in python for our system
- Can convert the graphically designed interface using QtDesigner to a python script easily.
- PyQt comes with all the advantages in both Qt and Python

2.12 Other Research works

2.12.1 Creating a Synthetic dataset using Conditional GANs

As there are no proper datasets which consists of vessel thermal images that is qualified for the task of re-identification we attempt to create a synthetic dataset. We decided to follow a image to image transition method and following approaches are considered. From the above method Pix2pix was considered to be the most appropriate

Pix2Pix[17]	Cycle GAN[47]	Thermal GAN[22]
Supervised	Unsupervised	Supervised
Paired images needed	Unpaired images can be used	Paired images needed +Thermal array
Only need to use input domain image to test the model	Only need to use input domain image to test the model	Need to use input domain image to test the model +Thermal array

Table 4: Considered image to image translation approaches

for the usecase.
As the initial results were not up to satisfactory several pre-processing techniques and hyper parameter changing methods have been carried out to improve the performance.

- Increase the number of epochs in the training process

- Using grayscale images to train the model instead of RGB images
- Augment the dataset by performing random jittering and brightness changes
- Using histogram equalization to change the contrast

2.12.2 Image super resolution

In the task of image super resolution, our goal is to upsample and enhance the resolution of a given low resolution image (LR) to create a high resolution image(HR). The behavior of optimization based image super resolution methods is mainly driven by the chosen objective function. To measure the performance of image resolution models, metrics like

- Peak Signal to Noise Ratio (PSNR)
- Structural Similarity Index (SSIM)
- Mean Opinion Score(MOS) are used

Image super resolution is popular in domains such as medical, media and surveillance, which is our use case. Here our goal was to extract the detected maritime vessels and objects by the object detection model and enhance the resolution of them for better re-identification by the re-identification model. Like in other sections of our pipeline, finding relevant images for training the super resolution model was challenging. Although there were many well known RGB domain super resolution datasets, there were no easily accessible publicly available thermal super resolution datasets. We were able to obtain a thermal super resolution dataset after contacting an author.

2.12.3 Conversion of one channel thermal image to three channel thermal image to improve object detection accuracy

Most of the object detectors like CenterNet, YOLOV4 etc. usually takes RGB images as inputs where it contains three layer. But as we are dealing with one channel thermal images where it contains one layer either we need to modify the network to have one layer for the input or we need to convert our images to have three layers. Otherwise

same features will be duplicated in each layer which would cause a waste in memory and computational power. As a solution to this and as a new research method to improve the detection accuracy we build up and tested a method to convert one channel thermal images to three channel thermal images.
The main idea behind the method is to process the original images with an image processing technique in each layer to highlight some useful features in the image and merge the layers to obtain a three channel thermal image. The below combinations of image processing techniques have been tested out.

- Canny edge detection + Thresholding + Orginal Image

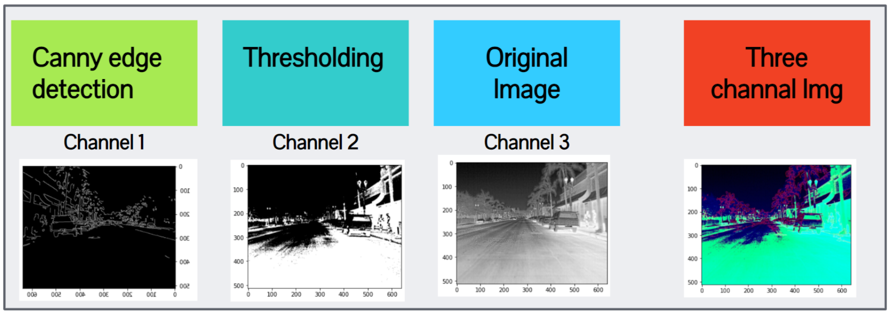


Figure 12: Techniques for each layer and the resultant image: Canny edge detection + Thresholding + Orginal Image

- Automatic canny edge detection + Adaptive gaussian thresholding + Histogram equalization

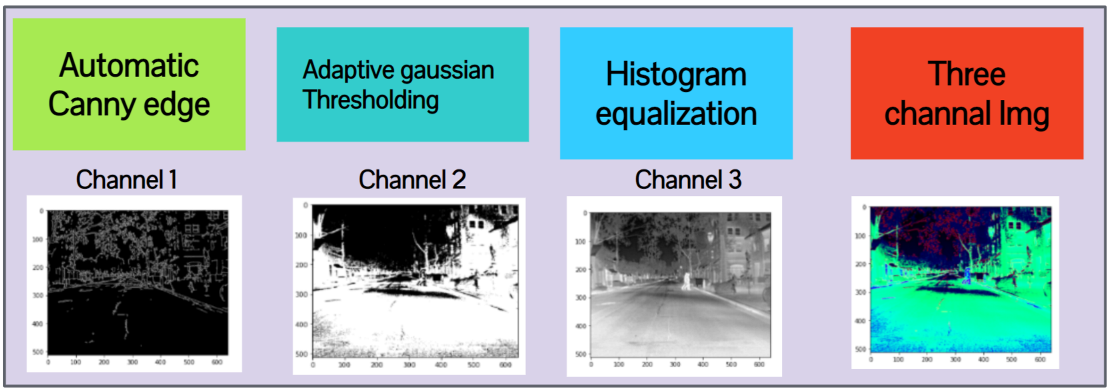


Figure 13: Techniques for each layer and the resultant image: Automatic canny edge detection + Adaptive gaussian thresholding + Histogram equalization

2.13 Evaluation Metrics

For the evaluation of the models, it's essential to choose a suiting evaluation metric for the problem, and these are essential to compare models based on their performance, speed, etc.

Mean Average Precision

In machine vision-based models, mAP is a very popular metric and it provides a precise evaluation of the models which consist of object detection as well as localization.

First, let us define the precision metric.

Precision provides how accurate our predictions are. It's the fraction of how many predictions are correct out of the number of total predictions.

$$Precision = \frac{TruePositives}{(TruePositives + FalsePositives)}$$

But if we only use this metric, we cannot determine the missed predictions as it only considers the predictions. Hence it will be a disaster for the model. Therefore, we need another metric to determine that. Therefore, we use recall for that.

$$Precision = \frac{TruePositives}{(TruePositives + FalseNegatives)}$$

Now we will discuss IOU, which is defined by the following equation,

$$Precision = \frac{Intersection}{Union}$$

In object detection and localization, we use bounding boxes and class labels. When training we define these two parameters and those are called ground truth values.

When a model predicted a bounding box with a class label, we say it is a true positive if that bounding box is closer to the ground truth bounding box and class label is correct. To check the closeness, we use IOU for that. Therefore, in object detection

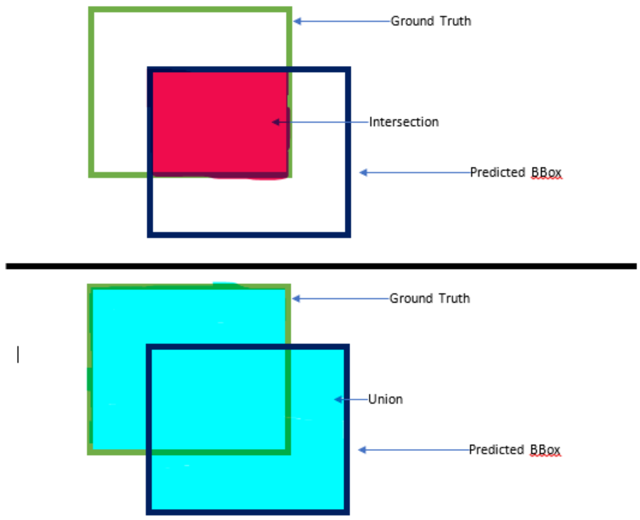


Figure 14: IoU Calculation Graphical Representation

problems, we calculate recall and precision according to the given IoU threshold.

For example, if IoU threshold is 0.6 and our prediction IoU is 0.7, we consider that as a True Positive. In the other way around, if the prediction is 0.3 we consider that as a False Positive.

It implies that according the IoU threshold, recall and precision will vary.

To calculate the mAP, we first must calculate the Average Precision, which is the area under the precision vs recall curve. As we know now, we can draw different curves for Precision vs Recall curves for different IoU thresholds. By taking the average of different AP, we can calculate the **mean Average Precision**.

$$mAP\ score = Expected\ Value(\Sigma AP)$$

Frames per Second

FPS is a critical evaluation metric in machine vision models when we must deal with video feeds. And it is more critical when we need that video feed in real-time.

It's defined such that number of frames that the model can give accurate predictions per second. If it matches with the FPS of the camera or if the camera FPS is lower than that of model FPS, then we can identify that model as a feasible model for our application.

Mean Square Error

Mean square error measures the average of squares of error. We can calculate MSE using following equation.

If Y_i is the observed value and $\hat{Y}_i$ is the predicted value.

$$MSE = \sum_{k=1}^{n} (Y_i - \hat{Y}_i)^2$$

Peak Signal To Noise Ratio (PSNR)

The PSNR block computes the peak signal-to-noise ratio, in decibels, between two images. This ratio is used as a quality measurement between the original and a compressed image. The higher the PSNR, the better the quality of the compressed, or reconstructed image.

$$PSNR = 10log_{10}(\frac{R^2}{MSE})$$

Cumulative Match Curve(CMC) Score

The cumulative match curve (CMC) is used as a measure of 1: m identification system performance. It judges the ranking capabilities of an identification system.CMC Rank n percentage metric shows how much of a percentage of the time correct prediction is ranked under n items.

Inter-frame Transformation Fidelity

Given a video I composed of Nf frames,ITF is expressed as the average inter-frame PSNR(Peak Signal to Noise Ratio).[14]

$$\mathrm{ITF} = \frac{1}{N_f - 1} \sum_{i=1}^{N_f - 1} \mathrm{PSNR}(t)$$

Inter-frame Similarity Index

Inter-frame Similarity Index (ISI) as the average of the SSIM between successive frames across the video.[14]

$$\mathrm{ISI} = \frac{1}{N_f - 1} \sum_{i=1}^{N_f - 1} \mathrm{SSIM}(t)$$

Average Speed

If a total of Np feature points are extracted in the video, the Average Speed(AvSpeed) metrics is given by[14]:

$$\mathrm{AvSpeed} = \frac{1}{N_p(N_f - 1)} \sum_{i=1}^{N_p} \sum_{t=1}^{N_f - 1} \|\dot{z}_i(t)\|_2$$

Average Percentage of Conserved Pixels

Given the original video I and a stabilized video ˜I, the Average Percentage of Conserved Pixels (AvPCP) could exeressed as the following ratio:

$$\text{AvPCP} = \frac{100}{N_f} \sum_{t=1}^{N_f} \frac{\text{res}(\tilde{I}_t)}{\text{res}(I_t)}$$

where res(.) is the resolution (in pixels)[14]

3 RESULTS

In this section we will provide our final year project progress that we were abled to achieve at the end of our project and also the results.

3.1 Thermal IR Camera

Figure 15: Video frame captured via the thermal camera

- We were able to create a camera library by analyzing its communication protocols and its packet structure. Using the library, we can send a dummy command to the camera to control it manually.
- This implementation can be used for the autonomous tracking of vessels.

3.1.1 Tests and Results

- We have done a range test with the thermal camera. Check the following results.

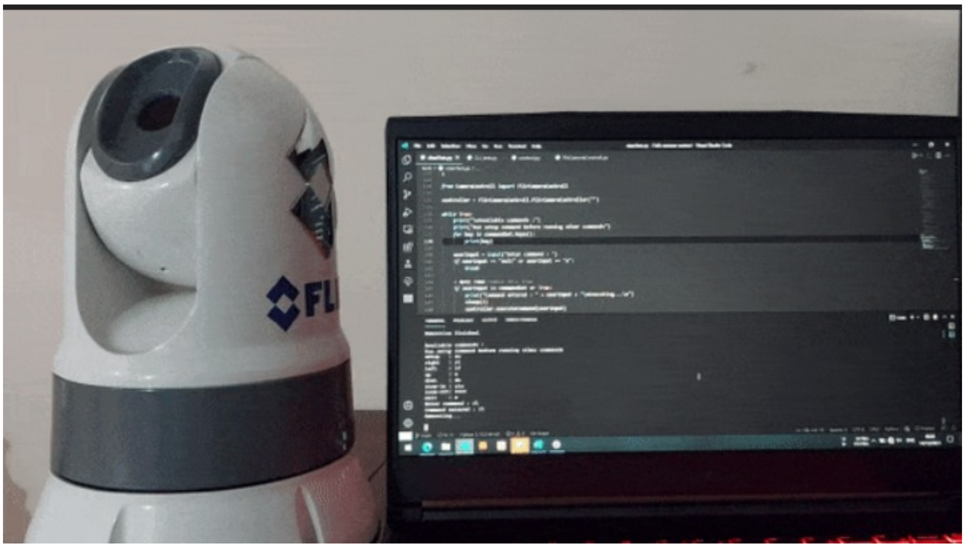
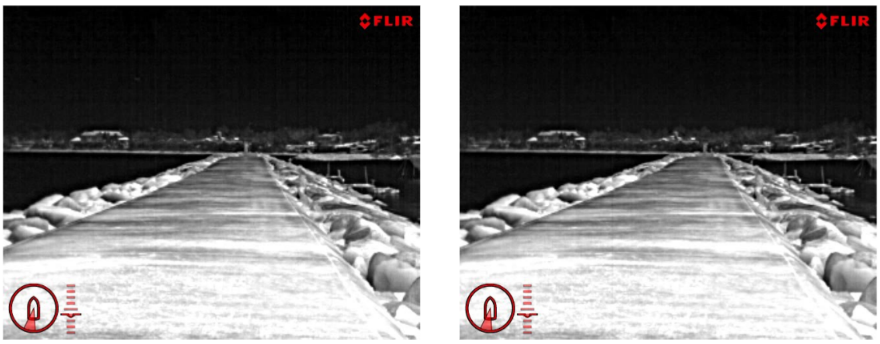


Figure 16: Camera controlled by python script

25m 50m 75m

Figure 17: Range Test with the thermal camera



100m 125m

Figure 18: Range Test with the thermal camera

- In the test we followed, we captured the human image and the distance and observed its visibility.
- Here we can identify that, the human image is detectable to the human eye at around 50m and more than that, the human is not recognizable.

- The next test we have done with the camera is comparison of the camera images at day time with the night time images. We have done this test to determine the properties of thermal images.
- Check the following results,

Figure 19: Day Image from the Thermal IR camera

Figure 20: Night Image from the Thermal IR camera

- From these results, we concluded that,
  - During the day time, temperature signature of the vessels are much visible to the thermal IR camera and it is harder to discriminate the humans onboard with the vessels.
  - During the night time, we can observe that humans onboard can be identified easily.

3.2 Self-Stabilization Platform

- Overall, we were able to create a final prototype of the platform according to the following anchor placements.

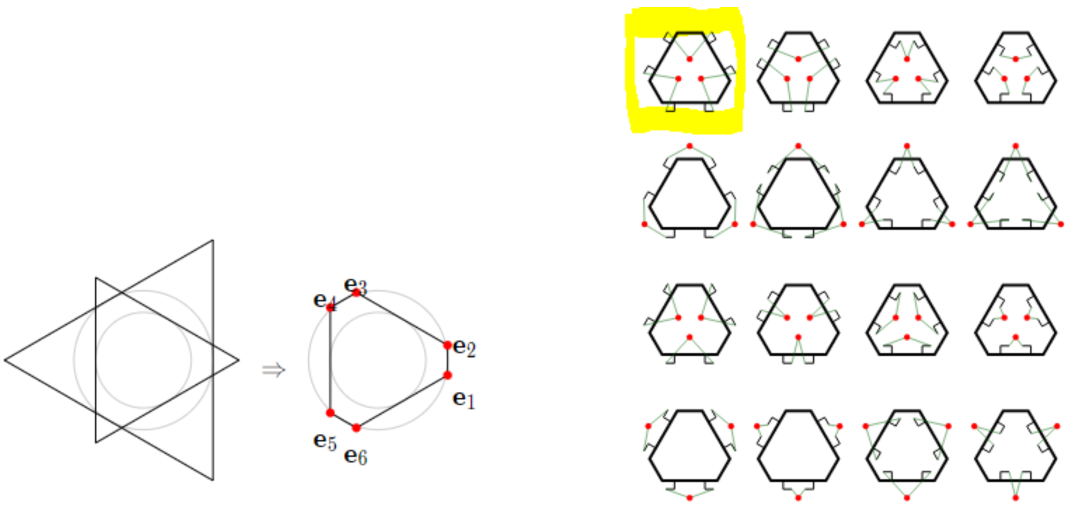


Figure 21: Anchor Placements[9]

- Here the e_k provides the placement of stage anchors and the highlighted one is used for our prototype designing and manufacturing.
- After considering above facts, through research, we designed the prototype through solid works for laser cutting purposes.(Not the Thermal IR Camera)
- The following provides the output of our prototype. For the joints, for 6 degrees of freedom functionality, we used ball joints, and to have a stabilized base, we used metal for the base, and for the stage, we used acrylic due to its lightness.

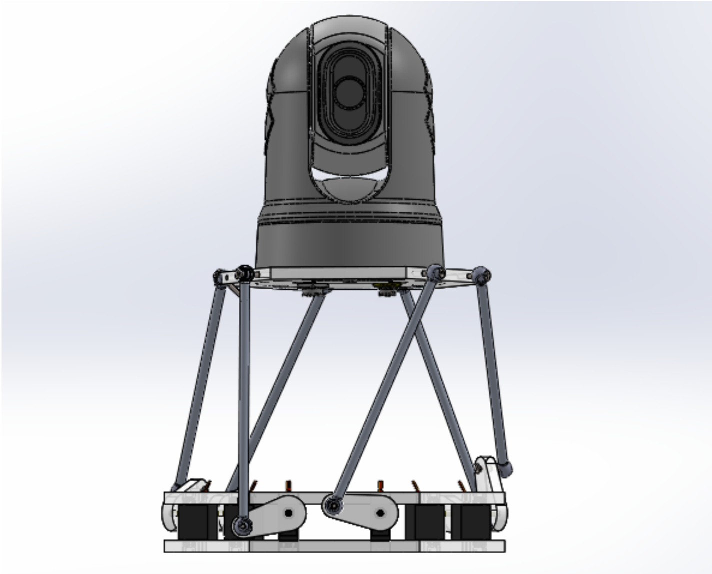


Figure 22: Solidworks design

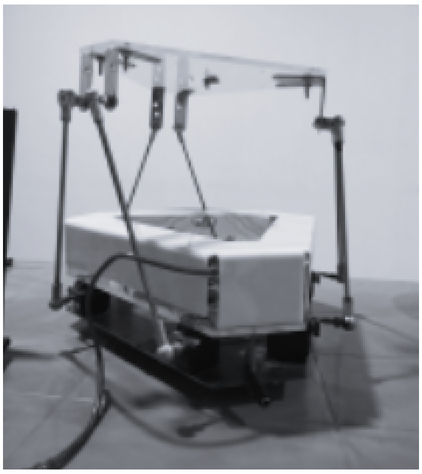


Figure 23: Final Prototype

- Using the this prototype platform, we programmed the platform to operate in two separate modes as we discussed before in the methodology section.

3.2.1 Manually Controllable Mode

We completed 100% of this part and following provides the results. In this experiment we provide a sine wave of amplitude 5 degrees and frequency 5 Hz as the input. Here in these graphs we can see the output of given by platform.

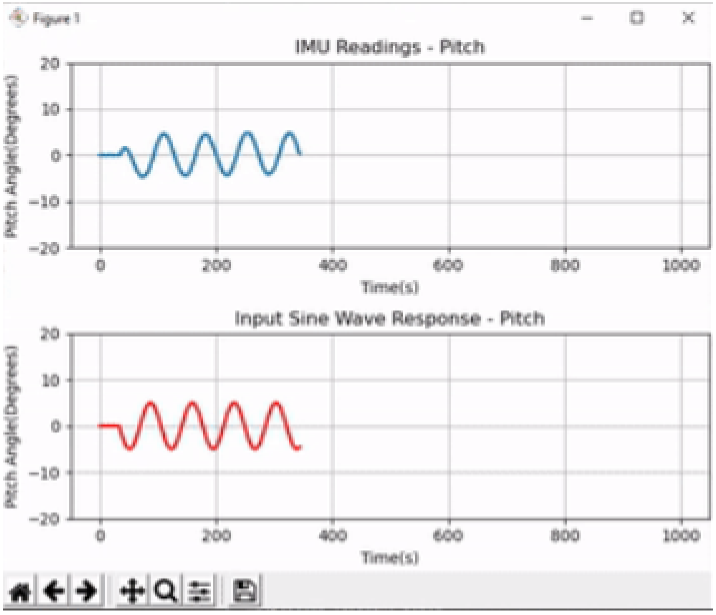


Figure 24: Results : Manual Controllability

3.2.2 Self Stabilization Mode

We completed 100% of this regard and here we evaluated the results using the following graphs which are given. In these graphs we obtain the orientation of the stage with respect to the global frame and orientation of base with respect to the global frame. Here we can see that although the orientation of the base is disturbed by the external motions, stage is abled to correct those errors and remained stable.

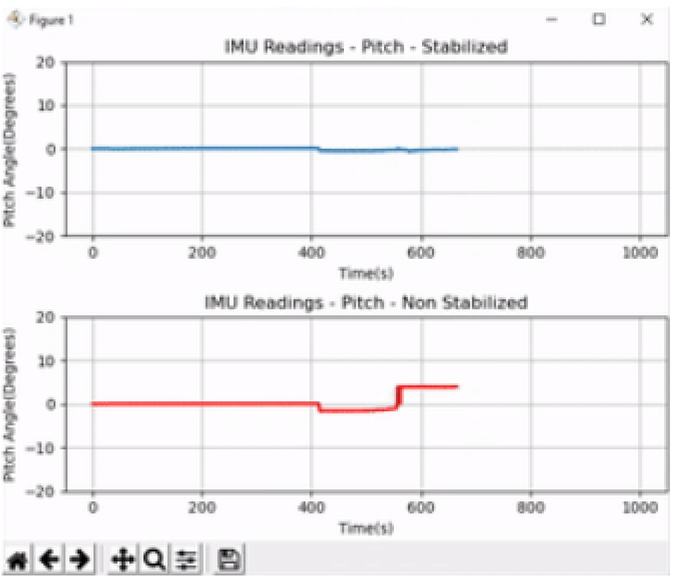


Figure 25: Results : Self Stabilization

3.3 Quantitative analysis of the Self Stabilzation Platform

- Original load : 3 kg (Tested)
- Maximum torque required by a single servo motor to operate : 0.295 Nm (30% of the stall torque of the Servo Motor)
- Steady state error measured : 0.01 - 0.03 rad.
- Workspace plot of the platform using geometrical constraints:

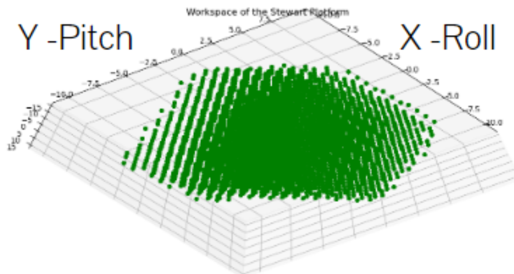


Figure 26: Workspace Plot

- Workspace:
  - Roll (rotation of the platform stage around x axis) : -6 degrees to 6 degrees.
  - Pitch (rotation of the platform stage around y axis) : -6 degrees to 5 degrees.
- Required Parameters for the system. Here we researched on the workspace requirements and we found out the following results. [31]

Motion:	Magnitude:
Surge	$\pm$ 0.23 m
Sway	$\pm$ 0.48 m
Heave	$\pm$ 1.22 m
Yaw	0°
Pitch	$\pm 3^\circ$
Roll	$\pm 10^\circ$

Figure 27: Caption

- This table was extracted from a master thesis [31] and they have done a extensive research and collected a dataset in rough sea conditions. Following table screenshot of Figure 20 provides the sea states,

WMO Sea State Code	Wave height	Characteristics
0	0 metres (0 ft)	Calm (glassy)
1	0 to 0.1 metres (0.00 to 0.33 ft)	Calm (rippled)
2	0.1 to 0.5 metres (3.9 in to 1 ft 7.7 in)	Smooth (wavelets)
3	0.5 to 1.25 metres (1 ft 8 in to 4 ft 1 in)	Slight
4	1.25 to 2.5 metres (4 ft 1 in to 8 ft 2 in)	Moderate
5	2.5 to 4 metres (8 ft 2 in to 13 ft 1 in)	Rough
6	4 to 6 metres (13 to 20 ft)	Very rough
7	6 to 9 metres (20 to 30 ft)	High
8	9 to 14 metres (30 to 46 ft)	Very high
9	Over 14 metres (46 ft)	Phenomenal

Figure 28: Sea States : Extracted from Wikipedia

- According to above table, for a rough sea situation of sea state 5, these parameters should be satisfied by the stabilization platform for a proper compensation.
- According to our requirement, we focus mainly on deep sea which is sea state 2 or 3. So, we can say that our workspace parameters are suits to the given constraints.
- Here the theoretical roll and angles were -10 to 10 degrees but due to the constraints in the servomotors and ball joints, the freeness of the movements were constraints upto some extent.
- Because of the above reason, Roll workspace dropped to 6 degrees and pitch also the same.

3.3.1 Software Stabilization

Following evaluation tables represent the metric values for the tested two software stabilization methods.

	ITF	ISI	AvgSpeed
Stable Video (Optimum)	43.44	0.95	0.05
Shaked Video (Sub-optimal)	34.10	0.82	2.12

Table 5: Optimum and Sub-optimal Reference Metric Values

	ITF	ISI	AvgSpeed	AvgPCP
Metric Values	34.54	0.85	1.60	100
Percentage of Stability Achieved in Each Metric	4.6%	24%	25%	100%

Table 6: After Applying Point Feature Mapping Software Stabilization- Metric Values and Achieved Stability Percentage of the Shaked Video

	ITF	ISI	AvgSpeed	AvgPCP
Metric Values	34.16	0.82	1.98	100
Percentage of Stability Achieved in Each Metric	0.64%	0%	6.8%	100%

Table 7: After Applying Mesh Flow Software Stabilization- Metric Values and Achieved Stability Percentage of the Shaked Video

When analysing the percentage of stability values, we can see the point feature mapping is good for proceed. But as it shows that percentage values are not good enough to compensate all the instabilities. Therefore, only the software stabilization can not stabilize the camera video feed. Implementation of a software stabilization method followed by hardware self stabilization platform is recommended.

3.4 Object Detection and Tracking

Following table represents the comparison of results that obtained by the selected object detector on several datasets.

Dataset	COCO (RGB)	MTIR (ours-ThermalIR)	FLIR (ThermalIR)
mAP	42.1	83.3	75.2

Table 8: Evaluation of CenterNet Object Detector on Datasets

Below image is displaying the Martime ThermalIR object detection and tracking output(tested using the ThermalIR dataset;MTIR).

Figure 29: Martime ThermalIR Object Detection and Tracking on UI

3.5 Action Detection

Since the action detection algorithm is evaluated by the detected objects in a image frame, the action detection accuracy in our method can be obtained by using the Mean Average Precision (mAP) score of the usued object detector. Following table presents the evaluation of CenterNet object detector on swimmers dataset,Singapore Maritime Dataset(SMD) and Maritime Thermal IR (MTIR)dataset.

Dataset	Swimmers (RGB)	MTIR (ours-ThermalIR)	SMD (RGB+NearIR)
mAP	54.9	83.3	60.7

Table 9: Evaluation of CenterNet Object Detector on Datasets

The swimmers dataset is containg low quality images and it did not have considerable amount of images per class label.Therefore,this mAP accuracy can be improved with a high quality images and with a balanced dataset.

Following images are displaying the final results that we achieved by our test images dataset.

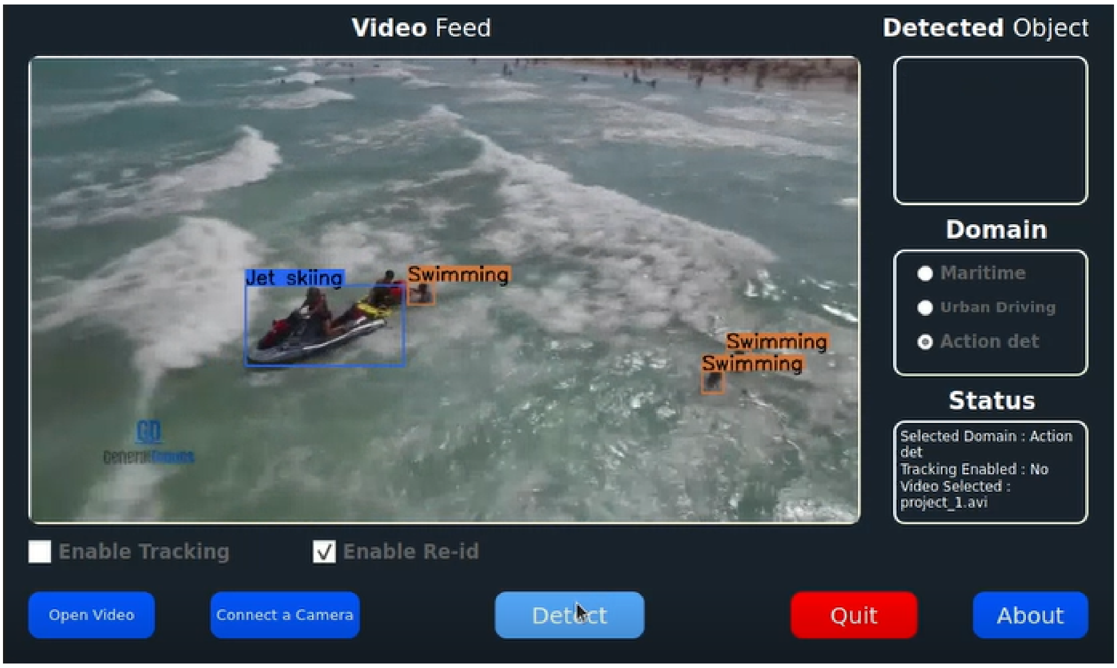


Figure 30: Jet-skiing and Swimming Action Detection and Localization on UI

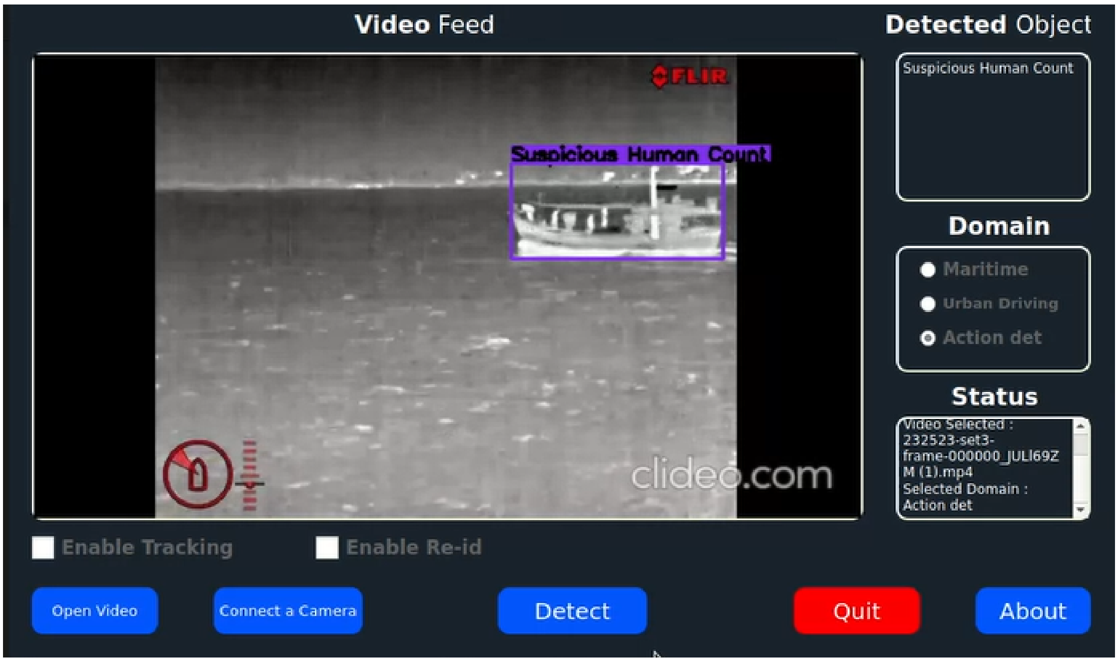


Figure 31: Unusual Human Count Action Detection and Localization on UI

NOTE : Maximum usual human count in a boat(a variable) is considered as 2 in the action detection algorithm. This variable can change appropriately. Due to the unavailability of real ”Unusual Human Count Action” images, we used the value 2 as for the testing purposes here.

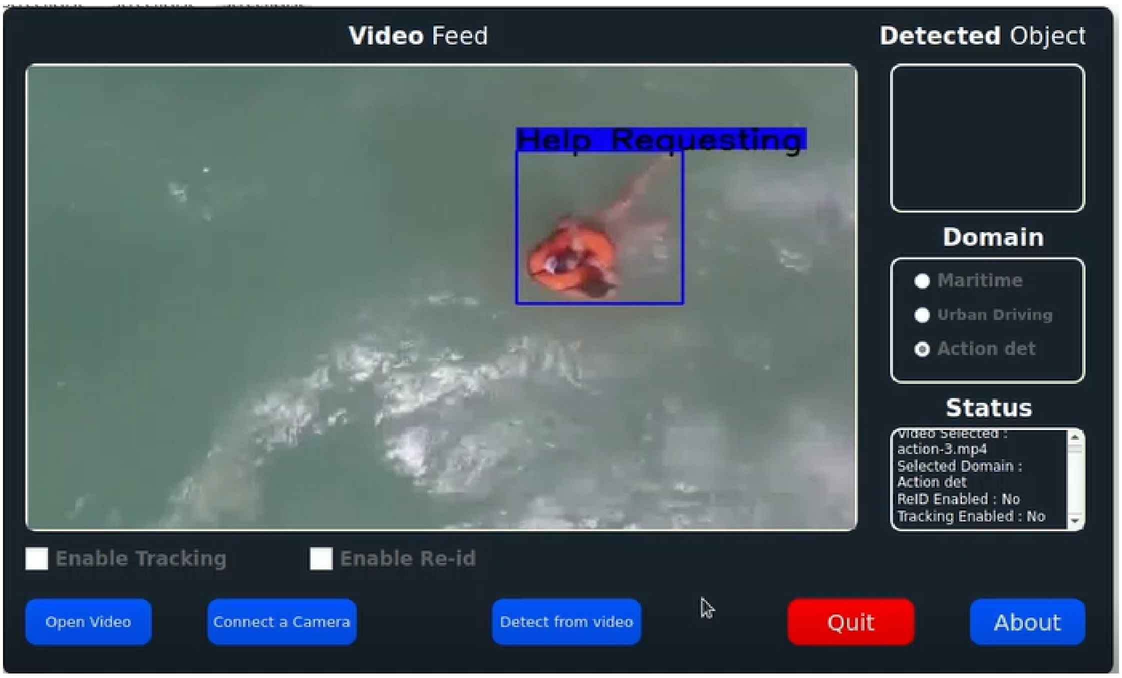


Figure 32: Help Requesting Action Detection and Localization on UI

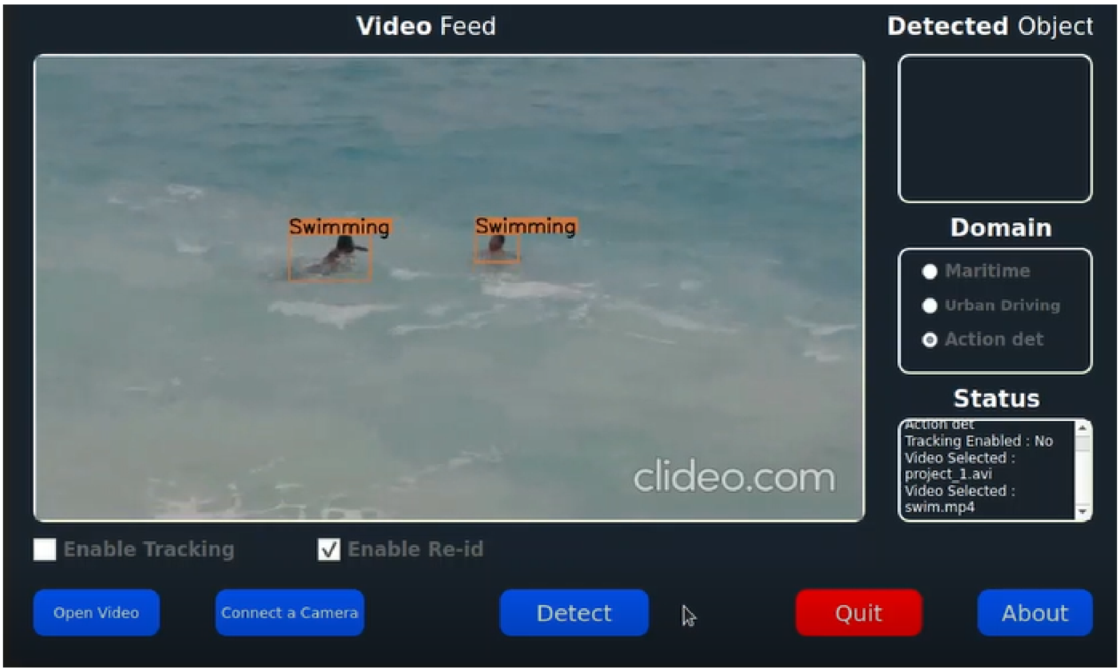


Figure 33: Swimming Action Detection and Localization on UI

3.6 Re-identification

In this section we will discuss the re-identification results we have achieved. After evaluation of the options as mentioned formerly, The strong baseline model was chosen to be used in the system.Due to the unavailability of thermal maritime vessel datasets, re-identification was done in thermal vehicle domain.

3.6.1 Strong baseline model results

First we were able to reproduce the model using the Veri776 RGB dataset and evaluate it to ensure the performance stated by the authors. As we are using a different dataset for the implementation of our task, the baseline model was tested using the RGB images of RGBNT100 (currently using the dataset). Also the model tested using thermal Ir images to verify the performance of it in the expected domain. The results obtained are shown in the 1st two columns of the provided table. As maritime re identification data are not currently available and could not be developed with the available resources, to train the re identification model with respect to our project's camera video feed and to demonstrate the functionality of the the complete system in re-identification , we created Vehicle Thermal IR (VTIR) data set with the FLIR M232, as mentioned prior in the report. The baseline model was trained on this dataset combined with the RGBNT100 data set and the performance was evaluated. The results obtained are displayed in the 3rd column of the table.

Dataset (domain)	RGBNT100 (Thermal IR)	Veri776 (RGB)	RGBNT100 + VTIR (ours) (Thermal IR)
mAP score	54.7 %	87.5 %	56.4%
CMC curve (Rank 1)	84.1 %	97.1 %	75.7%
CMC curve (Rank 5)	85.9%	98.8%	78.1%

Table 10: Results with different metrics

In the below figure you can observe how a detected car from a video feed is re-identified with the correct id by the complete system.

Figure 34: Re-identification results in the User Interface

3.7 Other research work

3.7.1 Creating a Synthetic dataset using Conditional GANs

The Pix2pix model performance for both Thermal IR and Near IR domains has tested using the RGBNT100[28] dataset

Below the original dataset means the dataset taken as it is without any pre-processing. Augmented dataset consists of a combination of actual and augmented images. When

Metrics	Epoch 200 - Original Dataset	Epoch 300 - Original Dataset	Epoch 200 - Augmented & Extended Dataset
MSE	1942.651624	2110.682486	2155.347749
RMSE	43.547668	45.342744	45.924001

Table 11: Results from Pix2Pix

evaluating the models using mean square error and root mean square error, the model trained for 200 epochs using the original dataset gave the best performance among others.
Also a comparison of the best model of Thermal IR domain and the best model of Near IR domain was done. When considering the above results it can be concluded that the Near IR model has outperformed the Thermal IR model.

Also the Near IR model has given better performance when inputting the images of

Metrics	Thermal IR (epoch 200)	Near IR
(epoch 200) MSE	1942.65162	889.84538
RMSE	43.547668	28.915584

Table 12: Comparison of Thermal IR vs Near IR

vehicles like lorries, buses and human images and converting them to Near IR images. Because of this model can be used for image to image translation tasks in other domains as well with domain adaptation techniques.

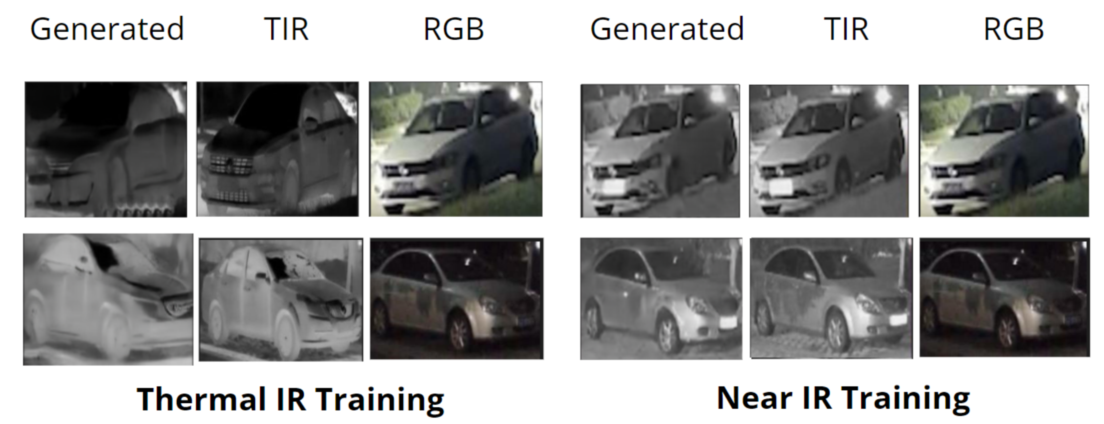


Figure 35: Near IR vs Thermal IR Training

3.7.2 Conversion of one channel thermal image to three channel thermal image to improve object detection accuracy

In order to test the performance of the method FLIR dataset has been converted using each approach. Then choosing the CenterNet as the object detector, the training process has been conducted for each modified dataset and the original dataset for different number of epochs. The following results has been obtained.

Number of epochs	Original FLIR	Approach 1	Approach 2
Epoch 100	0.731	0.713	0.683
Epoch 135	0.733	0.732	0.737
Epoch 200	0.752	0.745	0.740

Table 13: Comparison of the results for modified datasets and the original dataset

The highest accuracies are highlighted. It can be observed that the training performance of the original dataset has given the highest accuracies for most epochs. As there's no performance improvement for most epochs this research work has been discontinued.

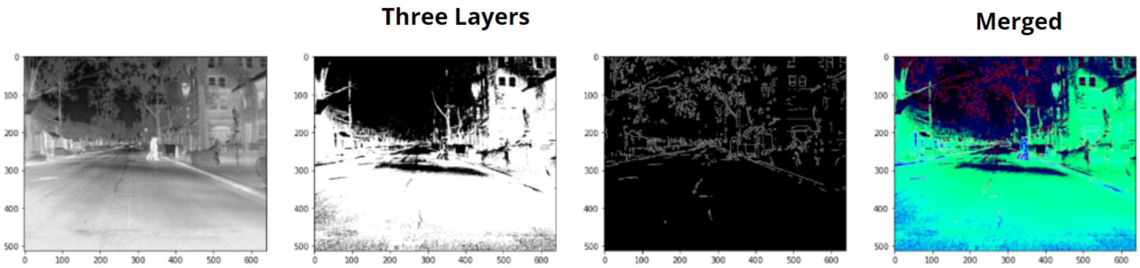


Figure 36: Three layers and the resultant image for approach 1- Canny edge detection + Thresholding + Orginal Image

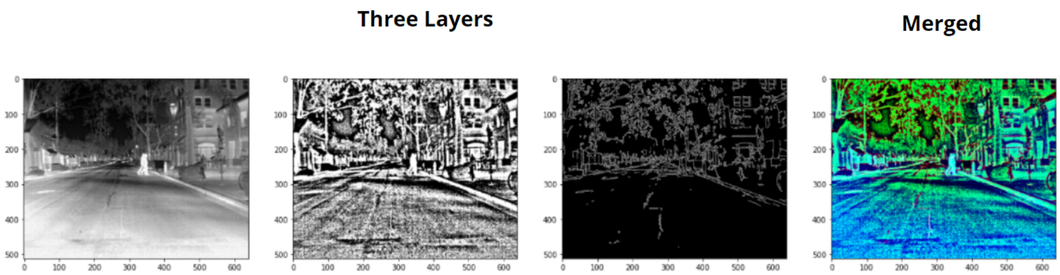


Figure 37: Three layers and the resultant image for approach 2- Automatic canny edge detection + Adaptive gaussian thresholding + Histogram equalization

3.7.3 Thermal image super resolution

Mainly 3 models were evaluated for this purpose.

Results with SWINIR[25]

For this a pretrained model was used for super resolving the low resolution thermal IR images.Images from RGBNT100 re-identification dataset were input to the model high resolution images were obtained.

Figure 38: Input and output images of SWIN-IR model sample 1

Figure 39: Input and output images of SWIN-IR model sample 2

The resulting images were highly smoothed and some features were not lost / correctly captured. The re-identification model performance on the super resolved images were tested out and comapared with the original images. The results are depicted in the folowing table.

Table 14: Input image

metric	Original Images	Super resolution images
mAP score	46.3%	38.4%
CMC rank 1	67.8%	55.9 %
CMC rank 5	70.0%	59.6 %

Overall re-identification performance was reduced using this approach with this model

Residual Dense Network (RDN) model[43]

A x2 scale RDN super resolution modal was trained on a thermal super resolution data set. After training a for 200 epochs the PSNR value reached was 20.9. This value is considerably lower than required to produce the required high resolution images.Sample images degraded in quality when fed to the model.

Figure 40: Input and output images of RDN model

Residual in Residual Dense Network (RRDN)model[38]

Similarly, a x2 scale RRDN super resolution modal was trained on the same thermal super resolution data set. After training a for 250 epochs the PSNR value reached was 20.69 . This value is also considerably lower than required to produce the required high resolution images.Quality of the output images were lower in quality than the input images.

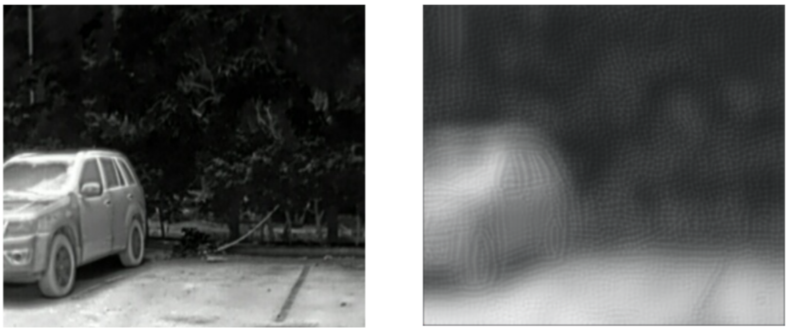


Figure 41: Input and output images of RRDN model

As all the approaches did not result suitable performance, further research was discontinued.

4 DISCUSSION AND CONCLUSION

Our aim in this project is to continue from the work done by the stage 1 team of the project to complete the development of an automated maritime surveillance system that is capable of object detection, tracking, re-identification, and activity detection paired with a stabilized camera platform. In tandem with this section of the project, we aim to develop an automated surveillance system.

To capture and provide the thermal IR video feed the system used the FLIR M232 thermal camera, which is a low-end maritime surveillance camera. Images from the formerly mentioned video feed would then be sent through the system created to detect, track, re-identify, and detect actions of maritime objects. Continuing from the stage 1 progress which included the system to detect and track maritime objects, we have developed the stabilization platform, achieved progress in utilizing the available thermal camera for the system, developed the activity detection, demonstrated feasibility in re-identification within the limitations of thermal IR domain, developed the model to convert RGB domain images to Near IR domain, achieved progress in developing the model to convert RGB domain images to thermal IR domain and created a set of annotated images required for training the models of the developing system. With the formerly mentioned achievements, we have made significant progress in achieving all the key deliverables of our project.

4.1 Principles, Relationships, and Generalizations Indicated by the Results

With the results that we have shown in this report and the progress, we have included we demonstrate that each of the components of the automated surveillance system that we are developing and the creation of the thermal annotated dataset is feasible.

The lack of a large thermal annotated dataset for training models and showing results was an issue we faced. We have demonstrated the ability to convert RGB domain data into Near IR data is feasible with the results shown in the report. With the data created in this manner, the full system could be developed in the Near IR domain which then can be trained to achieve desired results in the thermal IR domain as we believe. With the combination of data from SMD [32] and newly added annotations of the swimmers dataset made activity detection in the maritime domain more feasible because the lack of a dataset is the major challenge for the feasibility of the action detection component of the project.

We have developed full simulations of the stabilization platform. The results of the simulations demonstrate its feasibility to be used in our application.

The re-identification component of the project is currently developed on the vehicle re-identification domain due to the lack of a properly annotated thermal maritime re-identification dataset. But with the results we obtained in the thermal IR vehicle domain dataset, RGBNT100 [24], we demonstrate the possibility of the selected model[16] being used in the thermal IR domain.

4.2 Problems and Exceptions to the Generalizations

The main reason behind selecting the Pix2pix[17] model for image to image translation from RGB to Thermal IR is that even the authors have used the model to perform a task of translation Thermal IR images to RGB. As the RGB domain consists of more features than the Thermal imagery domain, translating from RGB to Thermal is comparatively easier than performing the other way around. But when training the model to do the translation task the model showed poor performance. In order to increase the results many pre-processing techniques and hyper parameters variations were tried, but those attempts were unsuccessful.

The major challenge to build the vessel re-identification model and activity recognition model is the lack of datasets to train them. The available SMD [32] and the FLIR dataset are limited to vessels and they have no images covering lots of maritime activities. In order to recognize broader set of maritime activities we need training images related to each activity. The plan to overcome the above challenge is to combine the swimmers dataset which has images covering lost maritime activities like swimming, surfing etc. with the Singapore maritime dataset [32]. Due to the unavailability of a dataset which qualified to train a model that can re-identify vessels, we decided to switch the domain and build a model that can re-identify vehicles. For that we get the chance to obtain the RGBNT100 multi spectral dataset [16] which consists of paired images of Thermal IR, Near IR and RGB.

The main challenge in the action detection part was the unavailability of a proper maritime domain dataset with maritime actions. And also FLIR M232 camera image resolution is also in a low level that makes hard to do the action detection using the state of the art methodology. Then as a solution we developed a new concept to detect actions using the object detection. This task would achieve more beneficial results to the community if there was any proper dataset. With the available resources we could able to develop the action detection which can be useful to detect actions is any domain if there is a suitable dataset. The provided solution in our work can detect actions in real-time even though the images are in low resolution.

4.3 Agreements/ Disagreements with previously published work

The results we obtained for re-identification in the RGB domain agree with the published results of the Strong baseline model [16]. The reproduced data provides the same accuracy on the same dataset.

The performance of the Pix2pix model [17] on the thermal vehicular dataset disagrees generally with the idea put forward by published work stating that the model provides good results on thermal IR data.

4.4 Theoretical and Practical Implications

There are some significant theoretical implications of this project, especially in the domain of activity detection using the object detection with a predefined set of rules. Although with the absence of a dataset, we implemented a logical approach by mapping the object detection and tracking results to solve the action detection problem. This is a comparatively easier approach than the state-of-the-art action detectors and the main advantage of using this approach is as we don't require a specific dataset containing suspicious activities which we cannot access from the security forces. Furthermore, inferencing this algorithm will be fast and time complexity will be reduced. This is more kind of a tailor-made approach for our problem and the only requirement is a dataset with human in the water, human onboard, human on earth, surf board, fishing rods likewise.

Furthermore, when we consider the object re-identification model, it has a huge practical implication as it also will be helpful to notify authorities about suspicious actions in an indirect way.

The importance of our project is that it holds a significant practical application in a country like Sri Lanka, as it is an Island covered with ocean and from our system, the security threat in our maritime borders will be covered autonomously. This project can be further developed if we can have the access to the suspicious activities data stored in the security forces which are related to our application and if we can have them then we can use a trainable action detection approach for that. Another significantly practical implication in our project is the use of deep learning models for widespread, automated surveillance. The system that we will develop during this project will prevent crimes that are happening and will affect hundreds of millions of human beings in a better way.

We have created a properly annotated maritime thermal IR dataset and hope to make it publicly available. This would be a great advantage for the future practical applications regarding the maritime domain,because there were no any publicly available maritime thermal IR dataset.

4.5 Overall Project Progress

In this subsection, we will discuss the overall work done by our team. As the deliverables for the project, we had to complete the Vessel Re-Identification model, fully annotated dataset creation in the maritime environment, maritime action detection, and self-stabilization platform. The following figure shows the overall completion of the project.

the stage I, the thermal maritime object detection model and object tracking algorithm were completed and we researched for the improvements that we can apply to that model and the algorithm. We considered several factors like speed, accuracy, and complexity as our system should be operated in a minimum resourceful area and should work in real-time with better accuracy. After considering several models like Yolo V4, Yolo V3 regarding the object detection models and Siamese Tracker, Deep Sort as tracking models, we concluded that Centernet is the best-suited model for object detection in real-time scenarios and SORT is the best-suited algorithm which has a less computational complexity for our application, and we started implementing our other models on top of these object detection and tracking.

If we consider the dataset creation, we were able to collect data during the day as well as at night in the maritime environment which includes 4 types of vessels and humans. Also, we created a Re-Identification dataset for vehicles including 5 vehicles, and re-annotated a complete dataset which is called the Swimmer's Dataset [30].

Regarding the Action Detection model, the major challenge we faced was the unavailability of a proper dataset to detect actions in the maritime environment in the Thermal IR domain. So, as discussed above we came up with a logical algorithm, to address this issue. And here we were able to complete and test the algorithm.

If we consider the Camera Platform, we were able to create two modes of operations for the platform and able to demonstrate.

Regarding the Re-Identification model, the absence of a proper dataset for the vessel re-identification was an issue that we faced as we did not get an opportunity to collect such type of dataset. So, we tested our vessel re-identification model for vehicles in Thermal Domain and were able to get better results.

Finally if we discuss our research paper, as we mentioned above we have done several types of research and finally, we decided to provide an application-based paper and were capable to draft it as well.

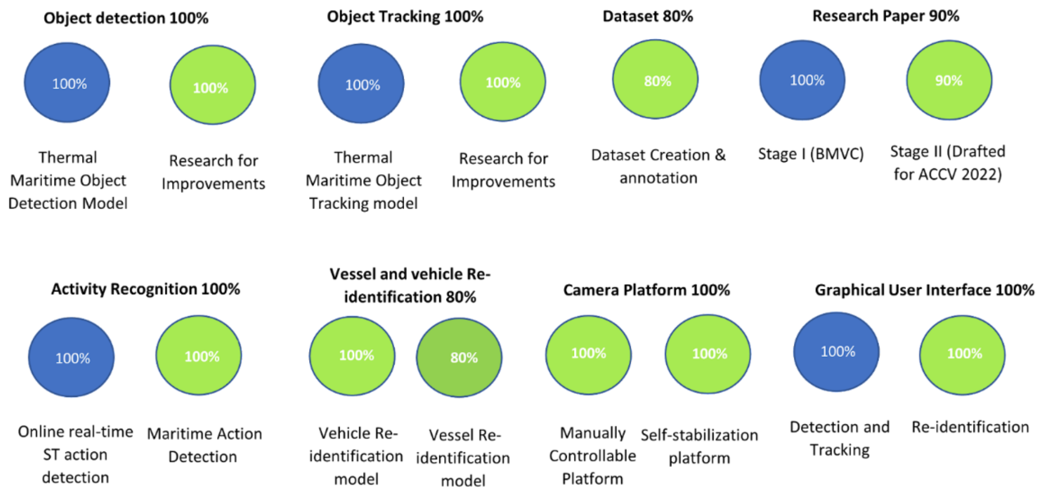


Figure 42: Overall Project Completion Diagram

4.6 Conclusion

According to our research and the results we gained, we can conclude that the implementation of a fully automated Maritime Surveillance System using thermal IR is possible although it is a challenging task. Since we use thermal IR, we can use this system in the nighttime as well as in the daytime.

This was a challenging task because although we have the models to train, we did not have the proper datasets for that. We were abled to create datasets by offshore in Beruwala and Mount Lavinia area but we did not get the opportunity to collect a dataset by mounting the camera on a ship/vessel. Finally we were managed to create a full pipeline of the system which is a fully machine vision network of Object detection through CenterNet, Object Tracking through SORT, Object Re-identification through Baseline Model for Re-identification, logical algorithm which we developed for action detection and the improved graphical user interface. On the other hand, we also abled to complete the Self Stabilization platform for the thermal IR camera and abled create several datasets related to the each application as well which includes vessels, cars, human onboard, human on earth etc.

Finally, we know that we have overcame lot of the challenges that we can endure and we were abled to finish our deliverables following Engineering practices, ethics and policies. Our work will be beneficiary for the ones who are interested in Thermal IR domain, Machine Vision domain and Maritime Environment, and Robotics and Automation.

References

[1] Jeonghyun Baek, Sungjun Hong, Jisu Kim, and Euntai Kim, *Efficient pedestrian detection at nighttime using a thermal camera*, Sensors **17** (2017), no. 8, 1850.

[2] Ganbayar Batchuluun, Dat Tien Nguyen, Tuyen Danh Pham, Chanhum Park, and Kang Ryoung Park, *Action recognition from thermal videos*, IEEE Access **7** (2019), 103893–103917.

[3] Luca Bertinetto, Jack Valmadre, Joao F Henriques, Andrea Vedaldi, and Philip HS Torr, *Fully-convolutional siamese networks for object tracking*, European conference on computer vision, Springer, 2016, pp. 850–865.

[4] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft, *Simple online and realtime tracking*, 2016 IEEE international conference on image processing (ICIP), IEEE, 2016, pp. 3464–3468.

[5] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao, *Yolov4: Optimal speed and accuracy of object detection*, arXiv e-prints (2020), arXiv–2004.

[6] Yunfei Cai, Shutao Zheng, Weitian Liu, Zhiyong Qu, and Junwei Han, *Model analysis and modified control method of ship-mounted stewart platforms for wave compensation*, IEEE Access **9** (2020), 4505–4517.

[7] Ruihang Chu, Yifan Sun, Yadong Li, Zheng Liu, Chi Zhang, and Yichen Wei, *Vehicle re-identification with viewpoint-aware metric learning*, Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 8282–8291.

[8] Kaiwen Duan, Song Bai, Lingxi Xie, Honggang Qi, Qingming Huang, and Qi Tian, *Centernet: Keypoint triplets for object detection*, Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 6569–6578.

[9] Robert Eisele, *Inverse kinematics*, `https://www.xarg.org/paper/inverse-kinematics-of-a-stewart-platform/`, Accessed: 2021-09-30.

[10] T Fossen, *Guidance and control of ocean vehicles. john willey & sons*, Inc., New York (1994).

[11] Thor I Fossen, *Handbook of marine craft hydrodynamics and motion control*, John Wiley & Sons, 2011.

[12] Jiyang Gao, Zhenheng Yang, and Ram Nevatia, *Red: Reinforced encoder-decoder networks for action anticipation*, arXiv preprint arXiv:1707.04818 (2017).

[13] Helmut Grabner, Michael Grabner, and Horst Bischof, *Real-time tracking via on-line boosting.*, Bmvc, vol. 1, Citeseer, 2006, p. 6.

[14] Wilko Guilluy, Azeddine Beghdadi, and Laurent Oudre, *A performance evaluation framework for video stabilization methods*, 2018 7th European Workshop on Visual Information Processing (EUVIP), IEEE, 2018, pp. 1–6.

[15] Xianpeng Zhu Aihua Zheng Hongchao Li, Chenglong Li and Bin Luo, *Multi-spectral vehicle re-identification: A challenge*, AAAI, February 2020.

[16] Su V Huynh, *A strong baseline for vehicle re-identification*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 4147–4154.

[17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, *Image-to-image translation with conditional adversarial networks*, Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134.

[18] Yang Jie, LilianAsimwe Leonidas, Farhan Mumtaz, and Munsif Ali, *Ship detection and tracking in inland waterways using improved yolov3 and deep sort*, Symmetry **13** (2021), no. 2, 308.

[19] Vijay John, Seiichi Mita, Zheng Liu, and Bin Qi, *Pedestrian detection in thermal images using adaptive fuzzy c-means clustering and convolutional neural networks*, 2015 14th IAPR International Conference on Machine Vision Applications (MVA), IEEE, 2015, pp. 246–249.

[20] Prashanth Kannadaguli, *Yolo v4 based human detection system using aerial thermal imaging for uav based surveillance applications*, 2020 International Conference on Decision Aid Sciences and Application (DASA), IEEE, 2020, pp. 1213–1219.

[21] Hilke Kieritz, Wolfgang Hubner, and Michael Arens, *Joint detection and online multi-object tracking*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 1459–1467.

[22] Vladimir V Kniaz, Vladimir A Knyaz, Jiri Hladuvka, Walter G Kropatsch, and Vladimir Mizginov, *Thermalgan: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset*, Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018, pp. 0–0.

[23] Hei Law and Jia Deng, *Cornernet: Detecting objects as paired keypoints*, Proceedings of the European conference on computer vision (ECCV), 2018, pp. 734–750.

[24] Hongchao Li, Chenglong Li, Xianpeng Zhu, Aihua Zheng, and Bin Luo, *Multi-spectral vehicle re-identification: A challenge*, Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 11345–11353.

[25] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte, *Swinir: Image restoration using swin transformer*, Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1833–1844.

[26] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg, *Ssd: Single shot multibox detector*, European conference on computer vision, Springer, 2016, pp. 21–37.

[27] Xinchen Liu, Wu Liu, Huadong Ma, and Huiyuan Fu, *Large-scale vehicle re-identification in urban surveillance videos*, 2016 IEEE international conference on multimedia and expo (ICME), IEEE, 2016, pp. 1–6.

[28] Xinjun Liu, Jinsong Wang, Tiemin Li, and Guanghong Duan, *Parallel mechanisms with two or three degrees of freedom*, Tsinghua science and technology **8** (2003), no. 1, 105–112.

[29] Fuchen Long, Ting Yao, Zhaofan Qiu, Xinmei Tian, Jiebo Luo, and Tao Mei, *Gaussian temporal awareness networks for action localization*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 344–353.

[30] Eleftherios Lygouras, Nicholas Santavas, Anastasios Taitzoglou, Konstantinos Tarchanidis, Athanasios Mitropoulos, and Antonios Gasteratos, *Unsupervised human detection with an embedded vision system on a fully autonomous uav for search and rescue operations*, Sensors **19** (2019), no. 16, 3542.

[31] Anders Lohmann Madsen and SG Kristensen, *Design of stewart platform for wave compensation*, Aalborg University (2012).

[32] Dilip K. Prasad, Deepu Rajan, Lily Rachmawati, Eshan Rajabally, and Chai Quek, *Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey*, IEEE Transactions on Intelligent Transportation Systems **18** (2017), no. 8, 1993–2016.

[33] Dalei Qiao, Guangzhong Liu, Feng Dong, She-Xiang Jiang, and Likun Dai, *Marine vessel re-identification: a large-scale dataset and global-and-local fusion-based discriminative feature learning*, IEEE Access **8** (2020), 27744–27756.

[34] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, *Faster r-cnn: Towards real-time object detection with region proposal networks*, Advances in neural information processing systems **28** (2015).

[35] Rafael E Rivadeneira, Angel D Sappa, and Boris Xavier Vintimilla, *Thermal image super-resolution: A novel architecture and dataset.*, 2020.

[36] Yunxiao Shan, Xiaomei Zhou, Shanghua Liu, Yunfei Zhang, and Kai Huang, *Siamfpn: A deep learning method for accurate and real-time maritime ship tracking*, IEEE Transactions on Circuits and Systems for Video Technology **31** (2020), no. 1, 315–325.

[37] Zygmunt L Szpak and Jules R Tapamo, *Maritime surveillance: Tracking ships inside a dynamic background using a fast level-set*, Expert systems with applications **38** (2011), no. 6, 6669–6680.

[38] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy, *Esrgan: Enhanced super-resolution generative adversarial networks*, Proceedings of the European conference on computer vision (ECCV) workshops, 2018, pp. 0–0.

[39] Nicolai Wojke, Alex Bewley, and Dietrich Paulus, *Simple online and realtime tracking with a deep association metric*, 2017 IEEE international conference on image processing (ICIP), IEEE, 2017, pp. 3645–3649.

[40] Fengwei Yu, Wenbo Li, Quanquan Li, Yu Liu, Xiaohua Shi, and Junjie Yan, *Poi: Multiple object tracking with high performance detection and appearance feature*, European Conference on Computer Vision, Springer, 2016, pp. 36–42.

[41] Lei Zhang, Helen Gray, Xujiong Ye, Lisa Collins, and Nigel Allinson, *Automatic individual pig detection and tracking in pig farms*, Sensors **19** (2019), no. 5, 1188.

[42] Yubo Zhang, Pavel Tokmakov, Martial Hebert, and Cordelia Schmid, *A structured model for action detection*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 9975–9984.

[43] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu, *Residual dense network for image super-resolution*, Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 2472–2481.

[44] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl, *Objects as points*, arXiv preprint arXiv:1904.07850 (2019).

[45] Yi Zhou and Ling Shao, *Aware attentive multi-view inference for vehicle re-identification*, Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6489–6498.

[46] Zongwei Zhou, Junliang Xing, Mengdan Zhang, and Weiming Hu, *Online multi-target tracking with tensor-based high-order graph matching*, 2018 24th International Conference on Pattern Recognition (ICPR), IEEE, 2018, pp. 1809–1814.

[47] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, *Unpaired image-to-image translation using cycle-consistent adversarial networks*, Proceedings of the IEEE international conference on computer vision, 2017, pp. 2223–2232.