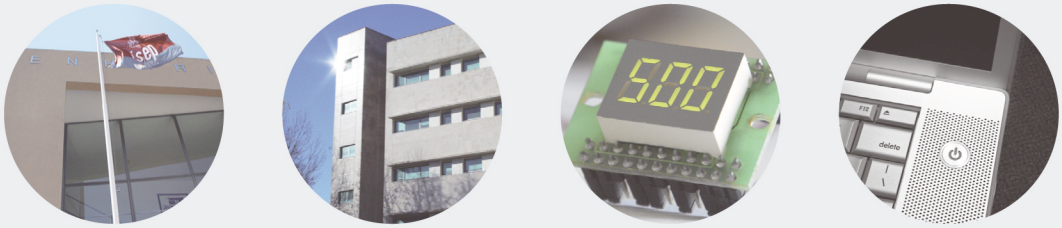
INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO
MESTRADO EM ENGENHARIA ELECTROTÉCNICA E DE COMPUTADORES



Object Detection and Tracking for ASV

DIMPI RAJUBHAI PATEL
novembro de 2020

POLITÉCNICO
DO PORTO

Instituto Superior de Engenharia do Porto

Object Detection and Tracking for ASV

Dimpi Rajubhai Patel

Masters in Electrical and Computer Engineering

Orientador: Professor Eduardo Alexandre Pereira Da Silva

November 24, 2020

Abstract

In this thesis automatic Object Detection system is presented. Object Detection is performed by different algorithms. As reading many literature we have observed that detecting objects in particular video sequence or by any surveillance cameras is a really challenging task in computer vision application because in sea the atmosphere affects a lot in the detection. Therefore we felt that there can be a wide range of possibilities are open in relation to detection. In order to improve the object detection, we developed image stabilization software on top of the image acquisition. First image stabilization has been performed over the raw data of ROAZ II. After achieving stabled video or images, object detection algorithm is performed using color based segmentation. Field tests have been performed with a data set from the ROAZ-II and during it shows the effectiveness of the approach. And system is able to achieve object detection in video or images with high accuracy.

Acknowledgement

I have taken efforts in this dissertation. However, it wouldn't have possible without many individuals. I would like to thank all of them .

I thank my God for providing me everything in this dissertation.

I would like to express the deepest appreciation to my Guide, my Professor Eduardo Alexandre Pereira Da Silva , who has attitude and substance of genius. Without his guidance and help this dissertation would not have been possible.

I would like to thank Professor André Dias for his kind guidance and support.

I would like to thank my course director Ms.Cecília Maria Reis for her kind guidance and support.

I would like to express my gratitude towards my family and friends for their kind cooperation and encouragement to achieve my goal.

Contents

List of Figures

List of Tables

List of Acronyms

ASV	***A**utonomous **S**urface **V**ehicles*
AIS	**A**utomatic **I**dentification **S**ystem
FAST	**F**eatures from **A**ccelerated **S**egment**T**est
HD	**H**igh **D**efinition
HSV	**H**ue **S**aturation **V**alue
ISEP	**I**nstituto **S**uperior de **E**ngenharia do **P**orto
LD	**L**ow **D**efinition
LIDAR	**L**ight **D**etection **A**nd **R**anging
LSA	**L**aboratório de **S**istema **A**utónomos
LWIR	**L**ong **W**ave **I**nfra **R**ed
MTT	**M**ulti **T**arget **T**racking
OTCD	**O**bject Level **T**racking and **C**hange **D**etection
RADAR	**R**adio **D**etection **A**nd **R**anging
RANSAC	**R**andom **S**ample **C**onsensus
RGB	**R**ed **G**reen **B**lue
SONAR	**S**ound **N**avigation **A**nd **R**anging
USV	***U**nmanned **S**urface **V**ehicles*
VTS	**V**essel **T**raffic **S**ervice

Chapter 1

Introduction

Autonomous Surface Vehicles (ASVs) are increasingly used in different applications for acquisition of field data or for monitoring water bodies and it is widely used for scientific and military reasons with guidance, navigation and control capabilities [14]. ASV is a robotic vehicle that stay on the water surface like sea or lake and records data across a different range of variables. Main advantage of ASV is the ability to operate without any human interaction. Considering the ability of an ASV to perform rescue missions,the thesis will address the development of a vision based approach for detecting and tracking targets on water. Object detection and tracking is an important and challenging task in Computer Vision which requires an ability to detect, recognize and track objects over a sequence of images called video.

1.1 Motivation

The rapid improvement in technology has made video acquisition sensor or devices better and it is low cost. Digital videos are a collection of sequential images with a constant time interval. So more information is present in the video about the object and background are changing with respect to time. Nowadays this autonomy system for an ASV can be more useful to detect and track vessels of a defined class while patrolling near fixed assets or the larger harbour area. After studying the many literature, i have seen that detecting and tracking of objects in particular video sequence or by any surveillance

camera is a really challenging task in computer vision application. Therefore i felt that there can be a wide range of research possibilities are open in relation to detection and tracking and it can be used in our robots to achieve good results.

1.2 Objective

This dissertation will address the development of a method that will contribute to the improvement of target detection method available in the ASV ROAZ. To accomplish we define some objectives:

- Evaluate the state of the art approaches to perform image stabilization, object detection and target tracking.

- Development of an algorithm to provide software based image stabilization.

- Development of an algorithm to provide software based object detection.

- Evaluate the results with a data set from the ASV ROAZ-II during some experimental field tests.

Chapter 2

Related work

In this chapter, a survey is made of similar technologies which has presented in research papers on detection and tracking.

In [2], Michael T. Wolfand team present an autonomy system for an ASV that detects and tracks vessels of a defined class while patrolling near fixed assets. They address two types of mission scenarios, each of which entails surveillance around a fixed asset. In each case, the system is trained to recognize type of boats of interest, referred to as targets because problem is ultimately one of multi target tracking (MTT). Missions were 1) Patrols large harbor region, 2) Patrols offshore region.

The algorithm receives information from an inertial navigation system and 6 cameras which are pointed 60 degree apart and having 5 degree of overlap between each adjacent pair. Image server captures image and stabilize the image. Contact server detects object of interest and calculate absolute bearing for each content. OTCD server (Object-Level Tracking and Change Detection) is finding target position and maintain database by true target and send downstream alert when new target appears or known target disappears. They successfully able to complete their task. For future work they want to train the contact detectors for handling multiple target classes, achieving not only detection but also classification from these algorithms.

In [3], author presents a graphical model for semantic segmentation of marine scenes was presented and applied to ASV obstacle-map estimation. To evaluate the perfor-

mance and analyze algorithm, from the literature survey they detail the HOG and distance classifier method to extract the fast and continuous obstacle from unmanned surface vehicle. The Histogram of oriented gradient use to find out the key feature of obstacle and normalize distance classifier find out the small and large object. This method extracts both small as well as large obstacle. Camera has placed on ASV to capture video on 360-degree angle from surrounding. Now this video has been converted into frame per second for further process. And this frame which is going to use as an input image for the process. Input to the system is an image in which there are classes like water,ground,sky,etc. Now prepossessing methodology point update of photo without changing the information. Now, semantic region divides the image classes into the category and divination carried in terms of structural feature of the image. Then the colour distribution on image is done. Water is considered as 1 and other obstacle is considered as 0. By using semantic segmentation, the obstacle detection has been done.The expected result shows the bounded box for obstacle having large and small obstacle.

In [4],Han Wang and team describe a method which aims at higher accuracy for the moving object detection and tracking in the open sea, and achievement successful target tracking range of within 500 meters in real time. Only one target is considered for tracking in this work. In this work, when the system for the ASV is started, it first detects all the possible objects on the sea and estimates their distances. Once a target object is selected, the tracking phase is enabled, meanwhile, the detection phase is stopped. If the tracked object is disappeared from the image or is canceled manually, the tracking stops, and the detection phase is enabled again. The pipeline of the proposed approach is first, the original high definition (HD) stereo images are resized to low definition (LD) for the sea surface plane estimation. Then the HD images are resized to medium definition (MD), on which the coarse object detection is done using the estimated sea surface plane on LD images. With the knowledge of the ROI for coarse object location as well as the sea surface plane, the location of objects are refined on the HD images. Finally, the tracking is performed on the HD images if the target is found. In the future work, they want to try to tackle the problem of multi-targets detection and tracking for the ASV.

In [5],Domenico Bloisi and team present an automatic maritime surveillance system. Boat detection is performed by means of an Haar-like classifier to obtain robustness

with respect to targets having very different size, reflections and wakes on the water surface, and apparently motionless boats anchored off the coast. Detection results are filtered over the time to reduce the false alarm rate. The system is able to provide the user a global view adding a visual dimension to AIS data. Camera with 26X optical zoom that can be moved by human user through control module and this module is also able to provide camera orientation and field of view. Detection module takes an input the current frame acquired by the camera. Output of this module is list of observations and each observation is a bounding box representing a detected boat. Output of visual tracking is set of visual tracks after filtering false positive. Only the tracks which presents a sufficient number of observations are considered of interest. In VTS system Radar and AIS data is merged to obtain graphical view. Validation module aims to give user an real time visual image for tracks. Data fusion between radar and visual tracks are performed on probabilistic. Like this way whole system works. The results on real data show the effectiveness of the proposed detection approach maintaining a 10-fps computational speed.

In [6],Nuno Pires and team present an innovative maritime surveillance system based on IR image processing. The ASV gives a visual dimension to detection. The proposed system works as below. Input has been taken from different sensors like Compass, Inertial Navigation Unit, GPS and cameras. Two types of cameras are used. One is fixed camera for 360-degree horizontal view and other one is rotational camera. Correction improves the quality of an image. Detection module extract relevant objects on image. Track module merge all relevant objects over multiple frame to useful track and eliminate false echo. Publishes track information to client system or to ASV's graphical user. It uses various output, particularly Smart Alarms by which the user can define his own alarm rule. Detection algorithm works like explain below. First module removes coast from the image then detects the horizon line and removes the sky. This removal allows ASV to carry out its object. Once relevant pixels are detected, each pixel belonging to an area is labialized as one detected object. Finally, the labelled area that don't touch the water surface or too small to be significant object are been discarded. After this tracking algorithm works as explain below. Track is a collection of individual detections representing an object along successive frame. Purpose of tracking is to monitor detected objects to build usable tracks and eliminate false echo. And we assume that detected objects which is not persistent in successive frames is false echo. The results of the detection are presented on a dedicated

user interface which emphasizes object crossing camera field of view.

In [7], Alfredo Martins and team present a set of field tests for detection of human in the water with an unmanned surface vehicle (USV) using infrared and color cameras. They did this experiments for the development of victim target tracking and obstacle avoidance for unmanned surface vehicles operating in marine search and rescue missions. For this work they have used the ROAZ unmanned surface vehicle equipped with a precision GPS system for localization and both visible spectrum and IR cameras to detect the target. USV requirement is to be able to detect victims in the water,this has two purposes: to contribute in the detection itself and provide rescue and for obstacle avoidance. First task in target detection was to detect Horizon in image and perform search below it. For that they have used edge detection coupled with a Hough transform for detecting large linear edges. The co linear segments are then grouped together and largest candidate group has been chosen. Now after this Horizon detection they have done Target detection. Histogram has been used to improve contrast. Threshold has been applied to image to to detect high temperature spot. High segmented blob below Horizon are segmented. Each blob centroid is possible target for the frame. In color image for target detection, They were using orange color life saving vests as a target. So color based image segmentation has been used and it performed in YUV color space with color limit. Detected target frame candidate was used n EKF based filter to eliminate reflection and false positive. The 3D target position has been determined with one image from infrared and other from color using standard pinhole camera model. From this method 3D positioning has been determined by triangulation and from this victim location has been achieved.

In [8], Alfredo Martins and team present an autonomous ground vehicle for outdoor exploration. The robot was designed in order to provide a versatile platform for multiple application robotics research in outdoor land scenarios and one of them was target tacking. So vision system is based on a pair of color cameras and infrared one. It's used in two type of functions, situational awareness directly in human super visioned task or in target detection and scene analysis image processing task such as intrusion detection and for navigation purposes. Image acquisition with color camera has been performed with an external synchronized trigger. It is set at a fixed rate to take images on left and right

for stereo processing and frame time stamping is needed for multi camera processing and future navigation based system . For each camera upon acquisition, global filtering has been operated on image to affect its properties. The processed image is then segmented according to suitable methods such as color segmentation, edge detectors or morphological operators. Then they apply region of interest or feature detectors to segmented images to identify relevant features or targets. Then they apply image processing in single frame pipeline basis or in a consecutive frame analysis. Target 3D positioning has been determined with images from the stereo color pair or one with infrared camera. They are using standard pinhole camera model. RANSAC algorithm has been applied to determine point position in the epipolar line. The vision system of robot assumes a sparse framework, so only target detected points are processed. For each corresponding pair of target image points on different cameras the relative 3D positioning is determined by triangulation. Target detection has been achieved through this. For the detected target positions (using the color camera stereo pair), robot trajectory and real target trajectory are indicated for a segment of the tracking maneuver when the robot is approaching the target and stopping afterwards.

Chapter 3

Fundamentals

3.1 Video Stabilization

Digital video stabilization is a process for reduce unwanted movement from a video stream. Generally, the processes of stabilization is composed by three phases namely: 1) Motion estimation 2) Motion smoothing 3) Motion compensation.

3.1.1 Method 1:FAST

Features from accelerated segment test (FAST) is a corner detection method which can be used to extract feature or interest points and then can be used to detect and map objects in many computer vision tasks. FAST is an algorithm proposed originally by Rosten and Drummond for identifying Interest points in an image. It is fast than many other well-known corner detection methods, such as SIFT, SUSAN and Harris [9].

Algorithm

We have used FAST algorithm for Video stabilization and implemented it in MATLAB.

•Read frames from Video file:

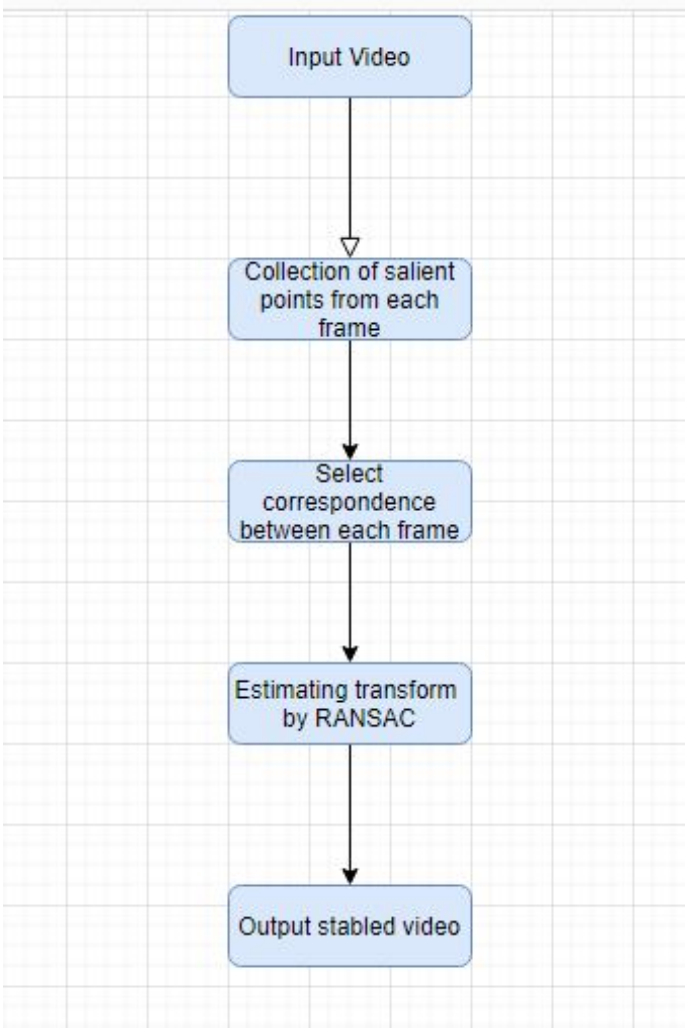


Figure 3.1: Stabilization algorithm: FAST

Reading them as an intensity images since color is not necessary for the stabilization algorithm.And also its showing total number of frames, height and width of video.

•Collection of salient point from each frame:

To see the pixel wise difference Red-Cyan color composite is performed.Goal is to determine the transformation that will correct the distortion between the two frames. To generate these correspondences, collection of points of interest from both frames has been performed.FAST the corner detection algorithm is used. To detect corner points we have to convert all the frames into grey scale images . Detected points from the consecutive frames are shown in figure.

•Select correspondence between points:

In this step, correspondences between the points derived in the previous step should be established.For each point in consecutive frame, we extract a Fast Retina Key point (FREAK) descriptor.The matching cost we use between points is the Hamming distance because FREAK descriptors are binary.Points in consecutive frames are matched putatively.Yellow lines are drawn between points because want to show the correspondences.Many of these correspondences are correct, but still there is significant number of outliers.

•Estimating transform from noisy correspondence:

Many of the point correspondences obtained in the previous step are identified with limited accuracy. To rectify this problem, Random Sample Consensus (RANSAC) algorithm is used. Because of limited accuracy we have observed that the inliers correspondences in the frame background are not aligned with foreground. The reason is the background features are far enough those act as if they were on an infinitely distant plane.We can assume that background plane is static and will not change certainly between consecutive frames.This transform is capturing the motion of the camera. Thus, correcting process will stabilize the video. As long as the motion of the camera between consecutive frame is minimize or the time of sampling the video is high enough, this condition is maintained.The result is calculated by projecting next onto previous via Sum of Absolute Differences between the two image frames. So, we can see that results are favorable.Transform approximation and smoothing performed on corrected frame sequences.

Further implementation of this we will be showing in chapter 4 The Proposed Solution.Use of FAST detection method shows improved result in term of stabilization.

•FAST detection technique is useful in enhancing the quality of video surveillance camera.

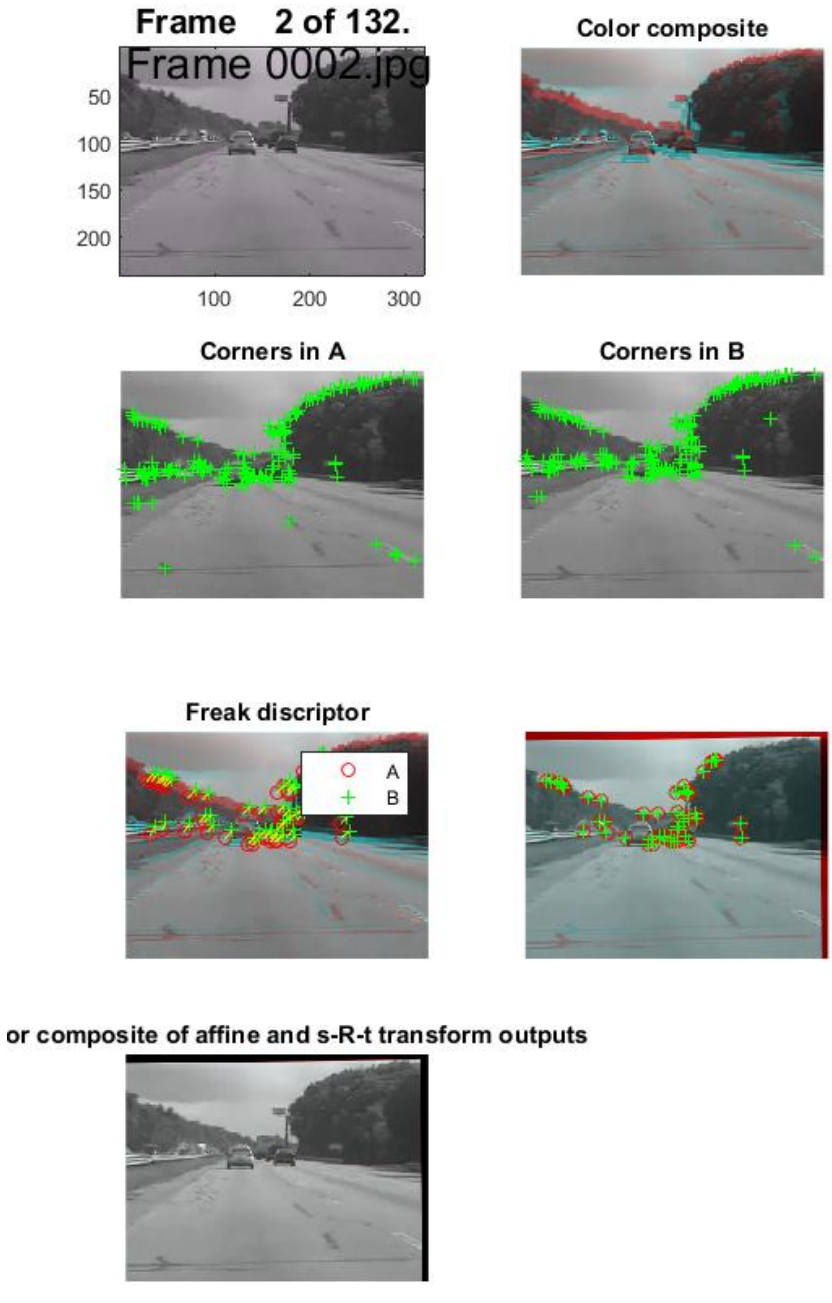


Figure 3.2: Stabilized image

3.1.2 Method 2: Lucas Kanade Optical flow:

Algorithm

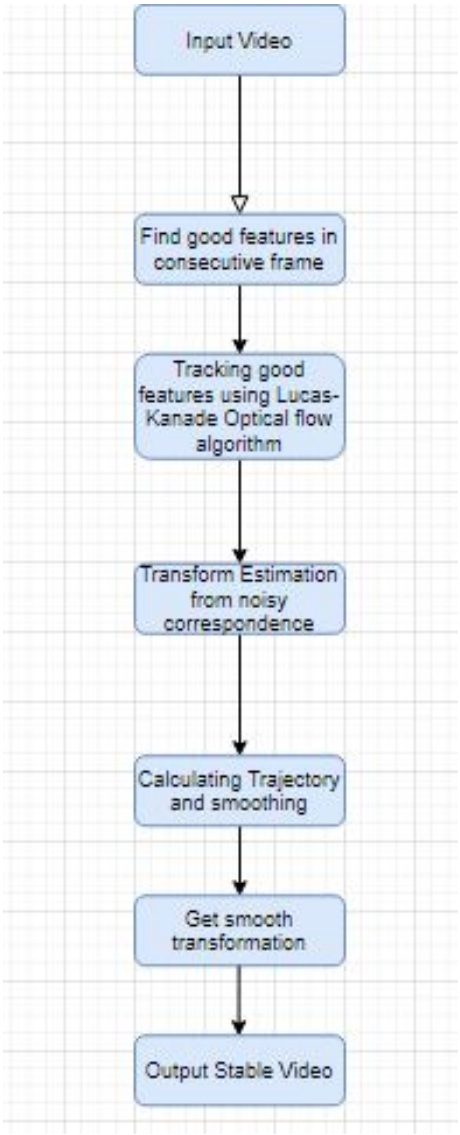


Figure 3.3: Stabilization algorithm: Lucas Kanade

•Read frames from Video file:

Reading frame as an intensity images since color is not necessary for the stabilization algorithm. And also its showing total number of frames, height and width of video and number of frames per second and approx it reads 30 fps.

•Finding good features in consecutive frames:

Reading next frame and converting it to gray. Opencv has a fast feature detector that detects features that are ideal for tracking. It is called goodFeaturesToTrack.

•Tracking good features using Lucas-Kanade Optical flow algorithm:

Once we have find good features in previous frame,we can track them in the next frame using an algorithm called Lucas-Kanade Optical Flow. This algorithm assumes that flow is essentially constant in local neighborhood pixels. Then we weed out bad matches to filter only valid points. We found the location of the features in the current frame, and we already knew the location of the features in the previous frame. So we can use these two sets of points to find the rigid (Euclidean) transformation that maps the previous frame to the current frame. This is done using the function estimateRigidTransform. In some cases if no transformation is found then the method uses the last known validated transform.We store this transformation and again move to next frame. And we get transformation for all the frames.

•Transform Estimation from noisy correspondence.

•Calculating Trajectory and smoothing it:

In this step we will add the motion between frames to get trajectory. It returns trajectory by making sum of differential motion dx,dy and da. As we have calculated trajectory,now we have three curves (dx,dy and da). Now we will smoothing this trajectory using moving average filter. The moving average filter is a simple Low Pass FIR (Finite Impulse Response) filter commonly used for regulating an array of sampled data/signal. It takes M samples of input at a time and takes the average of those to produce a single output point. As the length of the filter increases, the smoothness of the output increases, whereas the sharp modulations in the data are made increasingly blunt. So after applying this we get smooth trajectory.

•Getting smooth transformation:

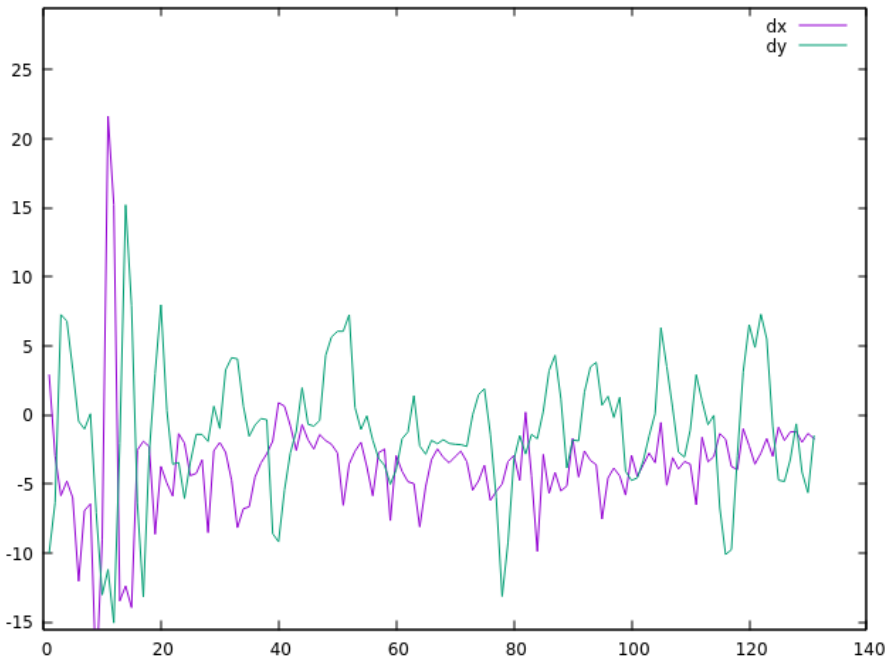
As we have calculated smooth trajectory, now we will use this to get smooth transformation and after applying this transformation to video we will be getting stabled video. Now we are doing this by getting difference between smoothed trajectory and the original trajectory and then will add this difference to original transforms.

•Stabilized video:

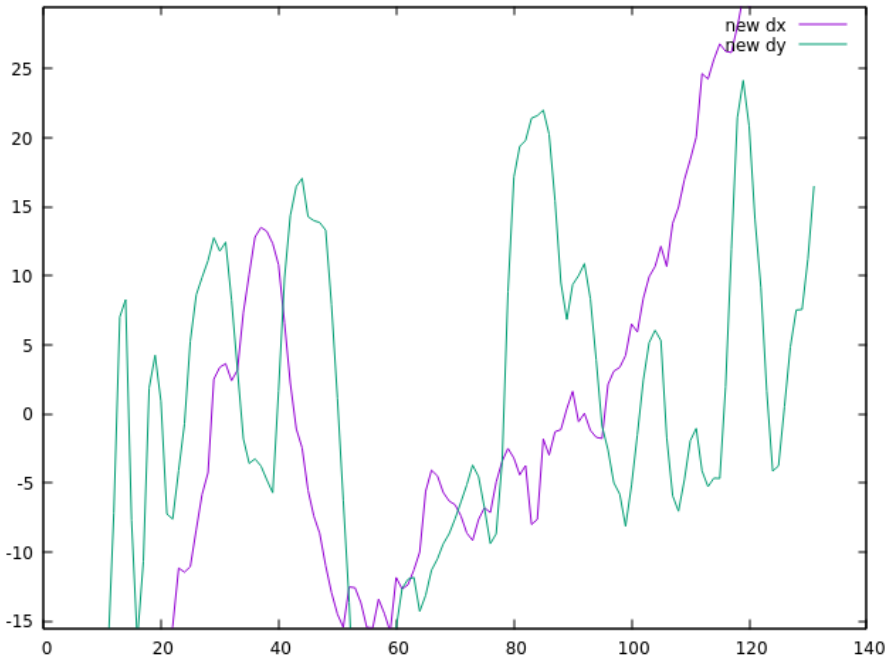
If we have motion (x,y,theta) corresponding transformation matrix is given by,

$$\begin{bmatrix} cos(theta) & -sin(theta) & x \\ sin(theta) & cos(theta) & y \end{bmatrix} \tag{3.1}$$

The method resetting stream to first frame. Reconstructing transformation matrix accordingly to new values. Applying affine wrapping to the given frame. Now we are showing the original and stabilized side by side for comparison. So as a output we are able to see that video is stabled.



(a) Previous transformation.



(b) New transformation.

Figure 3.4: Stabilized image graph by Lucas Kanade method

Above graph shows that in previous transformation has lot of jittery transforms during frames transformations and in new transformation it shows that transformation between frames have been improved and transformation are more smooth and less jittery. And this is achieved in OpenCV.

Based on the results, we are able to conclude that ,

• Both the methods have their own advantages.

• First method has been implemented by us in MatLab and it took more compilation time but user can see frame by frame implementation.

• Second method has been implemented by us in OpenCV and it took very less compilation time but user can see whole video stabilized not frame by frame.

3.2 Color Space: RGB and HSV

RGB is a color space with 3 dimensions: Red, Green and Blue. The RGB color space is the color space used by computers, graphics cards and monitors or LCDs and even by human eyes. Most digital images are stored as RGB images, and have to be converted to other color spaces. Color monitors display millions of colors by mixing different intensities of this primary colors. Range of intensity for each color on a scale is from 0 to 255. If all three color channels have a value zero that means that no light is emitted and the resulting color is black. And if all color channels have value 255 then resulting color is white. If we mix red and green, resulting color is yellow.

HSV is a color space with 3 channels: The Hue, The Saturation and The Value. HSV color space tells how our eye interprets/perceives scene to be. Hue channel represent color, The Saturation channel represents the amount of color. And the Value channel represents brightness of color. In Hue entire color has its one value. For an example, dark red also has the same value and light red also. The lightness or darkness of the color does not affect the Hue channel [12].

3.3 Object Detection

The requirement of performing object detection is present in several application areas like video surveillance, People Counting, Pedestrian detection, Tracking objects, Self-driving cars or Face detection, boat detection, buoy detection,etc. Object detection can be done in digital images and videos.Concept behind object detection is every object class has its own special features that helps in classifying the class – for example all circles are round. Object class detection uses the special features. For an example, when looking for circles, objects that are at a particular distance from a point (i.e. the center) are sought. Similarly, when looking for squares, objects that are perpendicular at corners and have equal side lengths are needed. A similar approach is used for face identification where eyes, nose, and lips can be found and features like skin color and distance between eyes can be found.

3.3.1 Techniques

Type of techniques to detect objects:

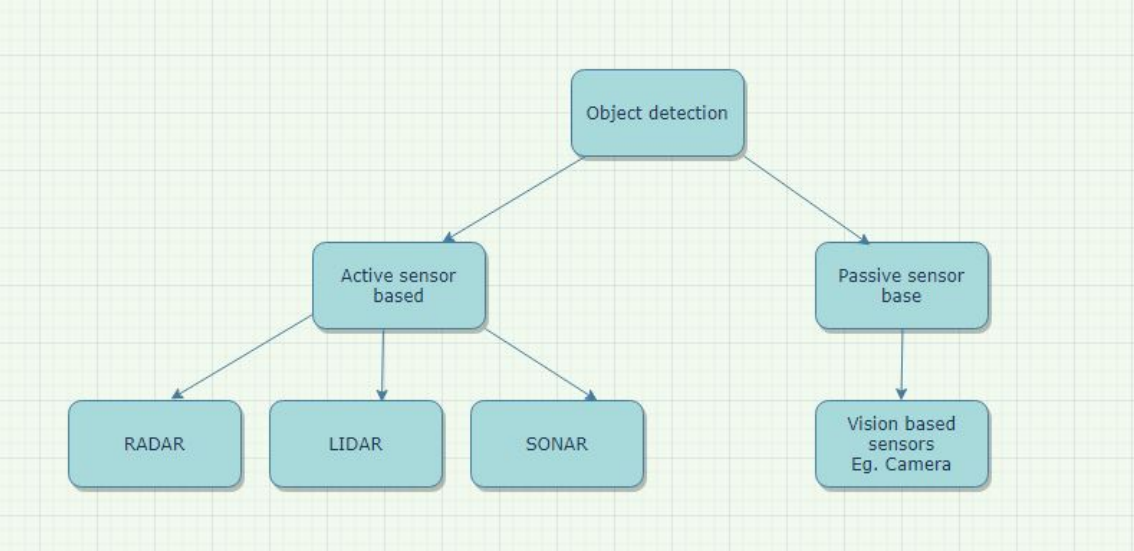


Figure 3.5: Detection Techniques

1. Active sensor based technique:

The sensor which provide their own energy source for illumination. The sensor emits radiation which is directed toward the target to be investigated. The radiation reflected from that target is detected and measured by the sensor. There are 3 type of active sensor through which object detection can be obtained: RADAR,LIDAR,SONAR [11].

a) RADAR:

Full form of RADAR is Radio Detection And Ranging. It is basically an electromagnetic system used to detect the location and distance of an object from the point where the RADAR is placed. It works by radiating energy into space and monitoring the echo or reflected signal from the objects. It operates in the UHF and microwave range. Radar systems are widely used in air traffic control, aircraft navigation, marine navigation. The RADAR system generally consists of a transmitter which produces an electromagnetic signal which is radiated into space by an antenna. When this signal strikes any object, it gets reflected or re radiated in many directions. This reflected or echo signal is received

by the radar antenna which delivers it to the receiver, where it is processed to determine the geographical statistics of the object. The range is determined by the calculating the time taken by the signal to travel from the RADAR to the target and back. The target's location is measured in angle, from the direction of maximum amplitude echo signal, the antenna points to. To measure range and location of moving objects, Doppler Effect is used.Most radar systems determine position in two dimensions:azimuth (compass bearing) and radius (distance). The display is in polar coordinates. A rotating antenna transmits RF pulses at defined intervals.

Figure 3.6: RADAR

b) LIDAR:

Full form of LIDAR is Light Detection And Ranging. It is a remote sensing method used for measuring exact distance of an object on earth's surface. LIDAR uses a pulsed laser to calculate an object's variable distances from the earth surface. These light pulses — put together with the information collected by the airborne system — generate accurate 3D information about the earth surface and the target object. LIDAR mainly consist of laser, scanner, specialized GPS receiver. There are two types of LIDAR : topographic and bathymetric. Topographic LIDAR typically uses a near-infrared laser to map the land, while bathymetric lidar uses water-penetrating green light to also measure seafloor and riverbed elevations.

Figure 3.7: LIDAR

c) SONAR:

Full form of SONAR is Sound Navigation And Ranging. SONAR is a way of communicating, navigating, and detecting objects by using sound propagation. It uses acoustical waves to sense the location of objects in the ocean. The simplest sonar devices send out a sound pulse from a transducer, and then precisely measure the time it takes for the sound pulses to be reflected back to the transducer. The distance to an object can be calculated using this time difference and the speed of sound in the water (approximately 1,500 meters per second). More sophisticated sonar systems can provide additional direction and range information. There are two types of sonar: active and passive. Passive sonar is a listening device only; sound waves produced by another source are received and changed into electrical signals for display on a monitor. Active sonar, on the other hand, sends out sound waves in pulses; scientists then measure the time it takes these pulses to travel through the water, reflect off of an object, and return to the ship.

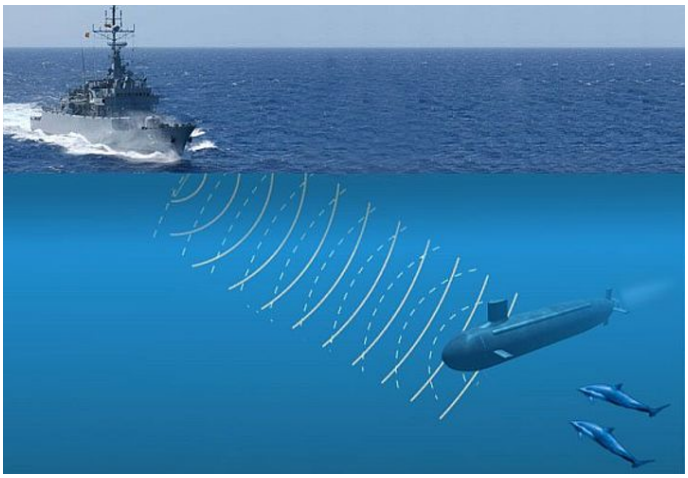


Figure 3.8: SONAR

3.4 Object Tracking

3.4.1 Introduction

Track is to follow the path of something. The goal of object tracking is segmenting a region of interest from a video frames and keeping track of its motion, positioning and occlusion over time. Object detection is the first step for tracking objects in successive video frames or in world. Object detection is performed to check existence of object in video frames. Object tracking is performed using monitoring objects' spatial and temporal changes during a video sequence, including its presence, position, size, shape, etc. Object tracking is used in several applications such as video surveillance, robot vision, traffic monitoring, etc. Object tracking is the process of tracking an initial set of detected objects.

3.5 Camera Model

We will be using camera model in this work so here I am explaining it in brief. Pinhole camera is most popular and simplest model. Pinhole camera basically is a box where light enters through small hole in front and produce an image opposite side. It is based on the fact that light rays travel in straight lines, so this workings of a camera will be able to reconstruct the 3D world from 2D images or vice versa. Mathematical model is as seen in figure.

In figure camera system C is represented by xc, yc,zc from origin O is camera center, the pinhole. Vector zc is called the viewing direction or optical axis. f is focal length. The point x is projection of X point in image plane.

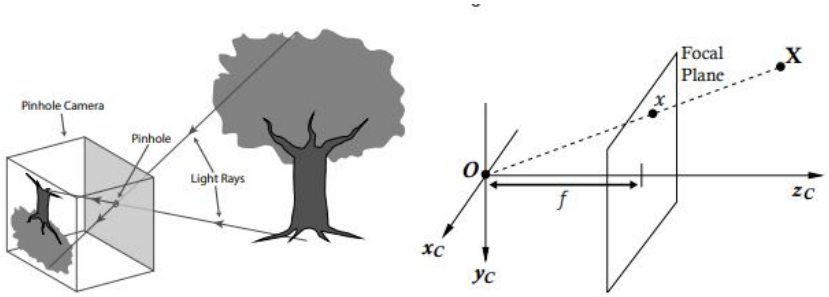


Figure 3.9: Pinhole Camera Model

3.5.1 Camera Intrinsic parameter

Images coordinates are measured in pixels, normally with the origin in the left upper corner. The focal plane in the pinhole camera model is embedded in $\mathrm{R}^3 soweneedtohaveamappingthattransl$

Considering pinhole camera model, image plane points relate to camera frame points by camera intrinsic parameters as given below:

$$K = \begin{bmatrix} f & 0 & cx \\ 0 & f & cy \\ 0 & 0 & 1 \end{bmatrix}$$

Where, K is intrinsic paraeters,

f is camera focal length,

cx and cy are the camera's optical centre.

cx= image width /2

cy = image hight/2

3.5.2 Camera Extrinsic parameter

The extrinsic parameters which denote the coordinate system transformations from 3D world coordinates to 2D camera coordinates. Equivalently, the extrinsic parameters define the position of the camera center and the camera's heading in world coordi-

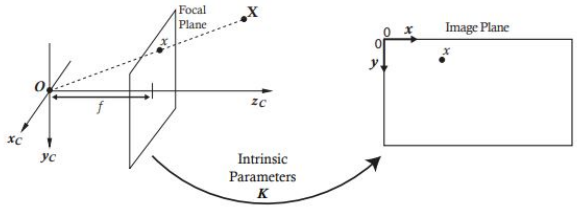


Figure 3.10: Representation of the transformation between the focal plane and image plane given by K

nates.It is used to describe the camera motion around a static scene, or vice versa, rigid motion of an object in front of a still camera. So the extrinsic parameters will be [13] .

Extrinsic Parameter= (R)T

where R is Rotational matrix,

T is translation vector

3.5.3 Camera Matrix

The projection matrix is used to convert from 3D read world coordinates to 2D image co ordinates or vice versa . If we know both intrinsic and extrinsic parameters than camera matrix can be calculated or we can convert 2D image coordinates into the world coordinates.

$$Pc = K \times Pw$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & cx \\ 0 & f & cy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r11 & r12 & r13 & t1 \\ r21 & r22 & r23 & t2 \\ r31 & r32 & r33 & t3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

When Pc is known it is possible to compute the image coordinates in pixels, of a 3D point in the world frame and vice-versa, hence the camera is considered calibrated.

3.5.4 Back-projection of a 2D image coordinates to 3D world coordinates

We will calculate real-world X Y Z coordinates from a given Image's projection points by following steps:

The pinhole camera model is define by following equation,

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & cx \\ 0 & f & cy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r11 & r12 & r13 & t1 \\ r21 & r22 & r23 & t2 \\ r31 & r32 & r33 & t3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Where, (X,Y,Z) are the coordinates of a 3D point in world frame,

(u,v) are the coordinates of a projection points in pixel,

A is camera matrix or intrinsic parameter,

(cx,cy) is a principal point that is image centre,

f is a focal length in pixel unit.

Now to get 3d points of world coordinate space ,

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \left(\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} A^{-}1 - t \right) R^{-}1$$

where, inverse A is inverse of intrinsic matrix,

Here ,

$$A = \begin{bmatrix} f & 0 & cx \\ 0 & f & cy \\ 0 & 0 & 1 \end{bmatrix}$$

t is translation vector,

inverse R is inverse of rotation matrix.

Algorithm

1. Get camera calibration done.

2. From this we will get intrinsic and extrinsic parameters

3. From extrinsic parameter obtaining rotational matrix and translation vector.

4. Finding inverse of intrinsic matrix and rotational matrix.

5. Getting u and v from image by finding the Centroids.

6. Now , substituting all the value in above equation we will get 3D points in world coordinates [14].

In detail we will explain it in chapter 5 The proposed solution.

Chapter 4

The Proposed Solution

For the Object detection and tracking we are proposing two solutions. But we have worked on detection in image or video. In future we can work ahead in detection in world.

A) Object Detection and Tracking in Image or Video.

B) Object Detection and Tracking in World.

4.1 Object Detection and Tracking in Image or Video

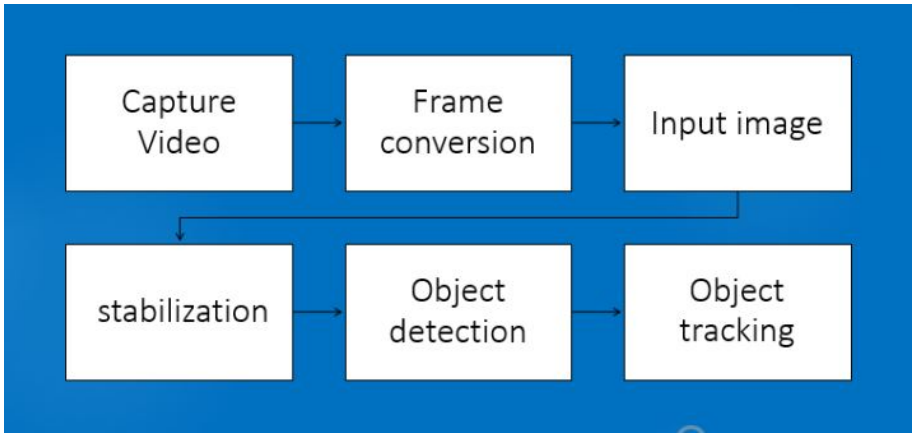


Figure 4.1: Detection and Tracking in Image or Video

As seen in above image, the algorithm is :

i) Read frames from video file.

ii) Stabilizing images or video.

iii) On stabilized images, objects has been detected.

iv) Tracking those detected objects in the video or images.

Now, we will explain all above points in detail.

i) Reading Video and converting video into image frames.And also its showing total number of frames, height and width of video.

ii)Stabilizing images or video.

There are two methods from which we have done video stabilization. One is with use of FAST algorithm and another one is with the use of Lucas-Kanade algorithm.

From above two methods we will be using FAST algorithm for the further processing.

iii) From stabilized video or images, object detection has been performed as below:

In our project we are going to use Passive sensor-based technique.

Passive sensor-based technique:

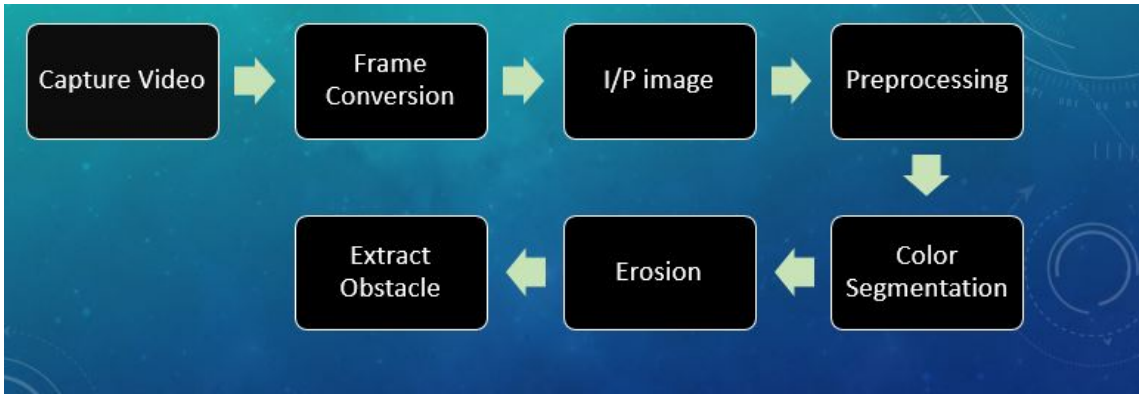
The sensors which does not have its own source of energy is known as Passive sensor. They have to take energy from external source. For an example, Vision based sensor. Camera has been used by vision-based sensor to capture the scene [11].

To detect object we are going to use HSV color segmentation technique.

4.1.1 Algorithm for detection

1.Reading Video and converting video into image frames.

2.Converted RGB image frames to HSV:



3.Color segmentation has been performed:

Finding which pixels represent the object. So, in this case H component always lies between 0.1 to 0.9 because 0.1 is starting value of the red color in HSV model and 0.9 is the last value which also represent the red color. So in between 0.1 to 0.9 we cover all the color. And Value component lies between 0 and 1. That is the ideal values.

4. Applying function erosion:

After applying this operation, we still get some noise. To remove or to get better result we used Erode function. Basically, morphological erosion removes small objects or noise so that only substantive objects remain.

5.After applying Erosion, we get the desired objects with very less noise.

6.And as we got objects in the scene, we made bounding box around detected objects.

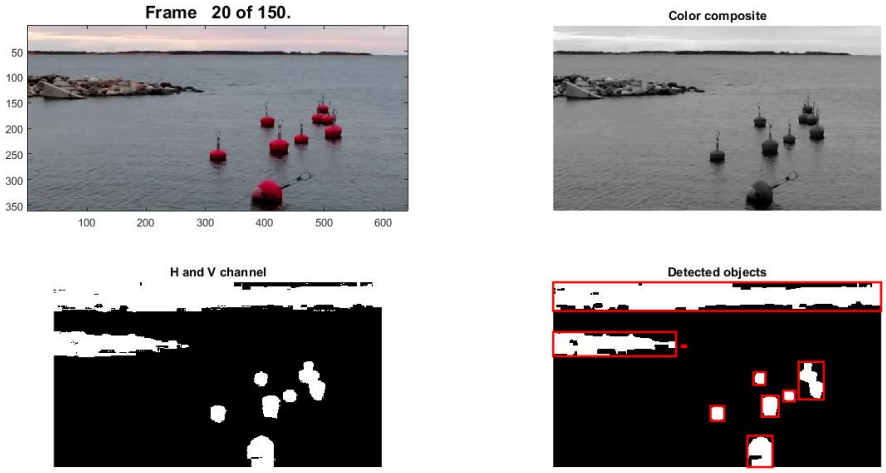


Figure 4.2: Detection

Chapter 5

Results

This results are achieved from the ROAZ II data set. So as you see in the images, the original image is in RGB format. As image has taken in sea so it has lot of waves and reflection so to detect the object in that format is bit difficult so we are using color based segmentation to get only objects. But before applying it we convert image into HSV format. After color based segmentation we apply erosion function to remove some noise and as you see in image only objects are left. After this we make a bounding box around the detected objects to get the perfect result. We are trying to detect both static and dynamic objects. Our algorithm detects objects below and above horizon line.

Now there are four terms use to see the accuracy of the algorithm and those are True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

True Positive (TP) : If it matches the same number of objects present in original data and in detection. For an example in figure 5.2 2 objects are present in the scene and it detects same number of objects.

True Negative (TN): If there are no objects in original data and also in detection it doesn't detect anything.

False Positive (FP): If the object is not present in original data but in detection it shows the object detected. For an example in figure 5.4 only 1 object is there in scene but it detects 6 objects.

False Negative (FN): If the object is present in original data but in detection it doesn't shows any detection. For an example in figure 5.6 there are 2 objects present in scene but it detects only 1 object.

We have processed algorithm in 2 parts: First we perform image stabilization and then we have achieved detection. First we have converted video into RGB format as original video is taken from thermographic camera. To convert this first we have converted thermographic image to HSV using cat(1,single(image)./255) and then we have converted HSV image to RGB. Then we are processing the intensity value of the images since color is not necessary for stabilization algorithm. Then to see the pixel wise difference we have performed Red Cyan color composite. Then we use FAST corner detection algorithm to detect corner points and for that we have converted frames into gray scale frames. For each point in consecutive frame we extract FREAK descriptor. Points in consecutive frames are matched and most of them are correct but still to remove outliers RANSAC algorithm we have used. We assume that background plane is static and will not change certainly between two consecutive frames. So correcting this process , stabilize the video. So,we can see that results are favorable. Transform approximation and smoothing has performed on all the corrected frame sequence. Below figure shows stabilization .

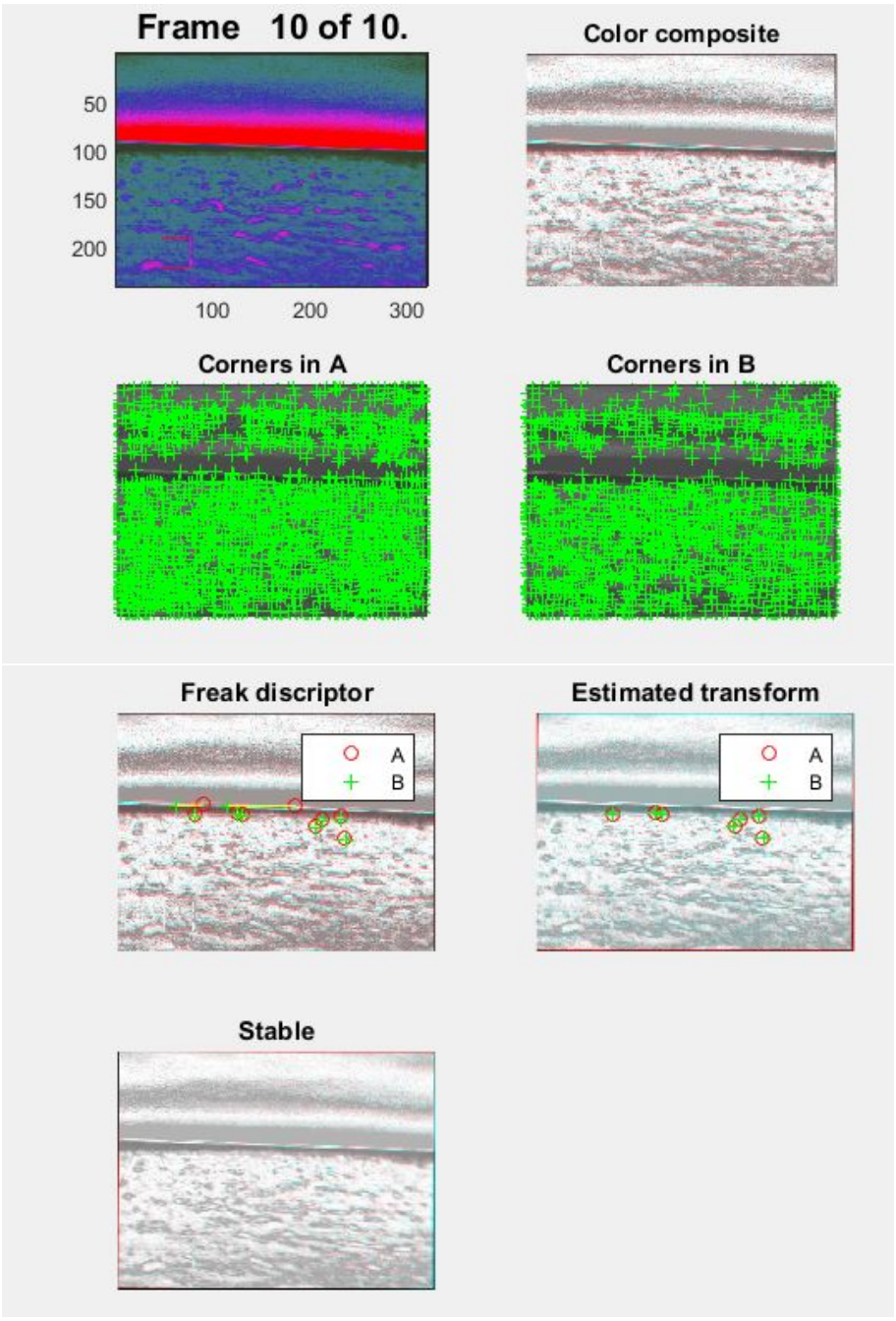


Figure 5.1: Stabilization

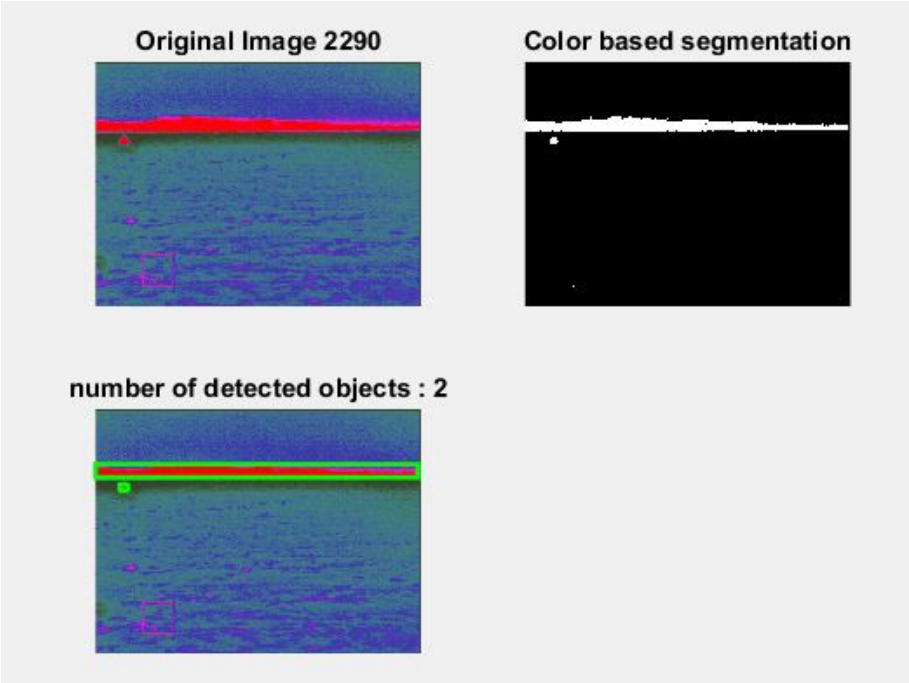


Figure 5.2: Detection in image

As we can see in figure 5.2, original image is in RGB format. Then we convert it into HSV format. Now to detect objects we apply color based segmentation . So only objects are left in image. Now to remove unnecessary noise we apply erosion function. After applying this we only get objects. And then we make bounding box around it. Detection has been performed in approx 0.421 sec. As we can see in image, objects which are present in original data has been correctly detected so number of True Positive is 1 and there are no False Positive and False Negative.

Here in figure 5.3, number of true positive frame is one and average CPU time is 0.405 sec.

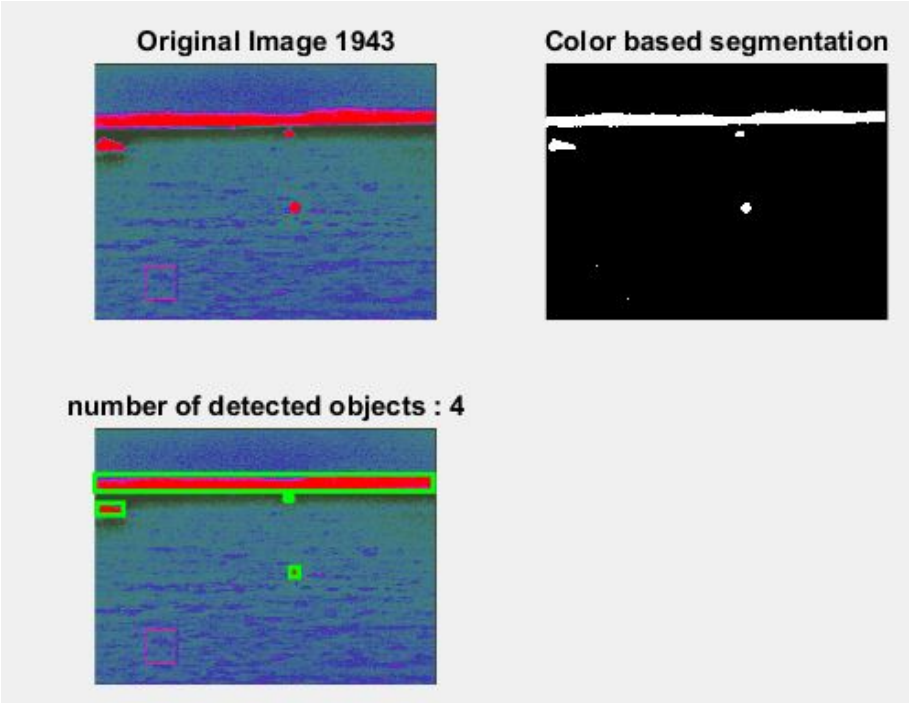


Figure 5.3: Detection in image

As we see in figure 5.4, in original image only one horizon is present but in detection it shows six objects has been detected which is wrong. So here we have one False Positive frame. And average CPU time is 0.370 sec.

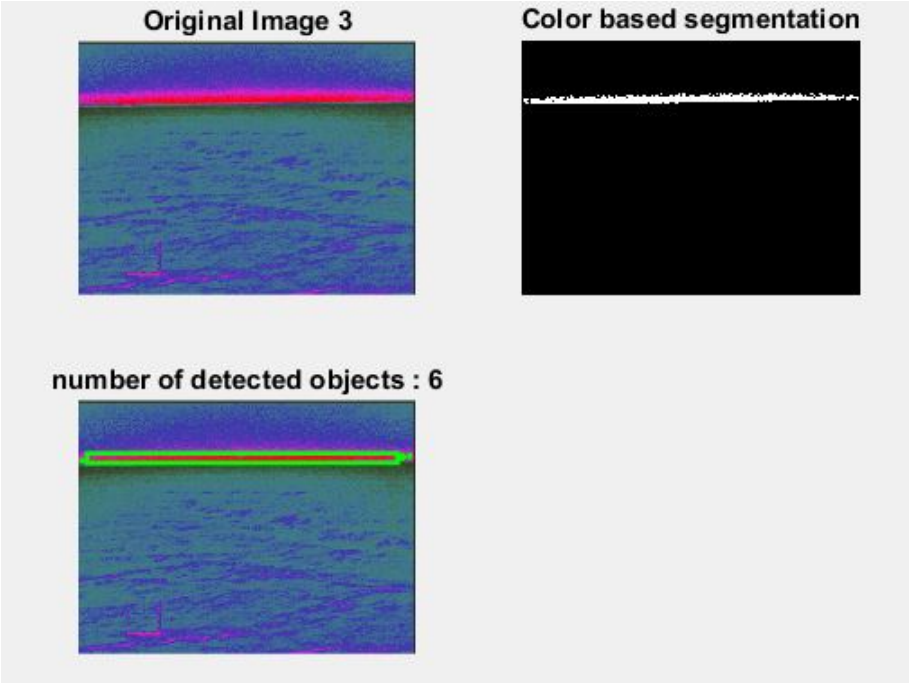


Figure 5.4: Detection in image

As we see in figure 5.5, original video is in RGB format. Then we convert all frames to HSV format. Now to detect objects we apply color based segmentation . So only objects are left in all the frames. Now to remove unnecessary noise we apply erosion function. After applying this we only get objects. And then we make bounding box around it. Here we have done this process on 10 frames and number of True Positive detection is 9 frames and there is one False Positive frame has been detected. As you see in last frame in horizon it count one more object in real which is not present.

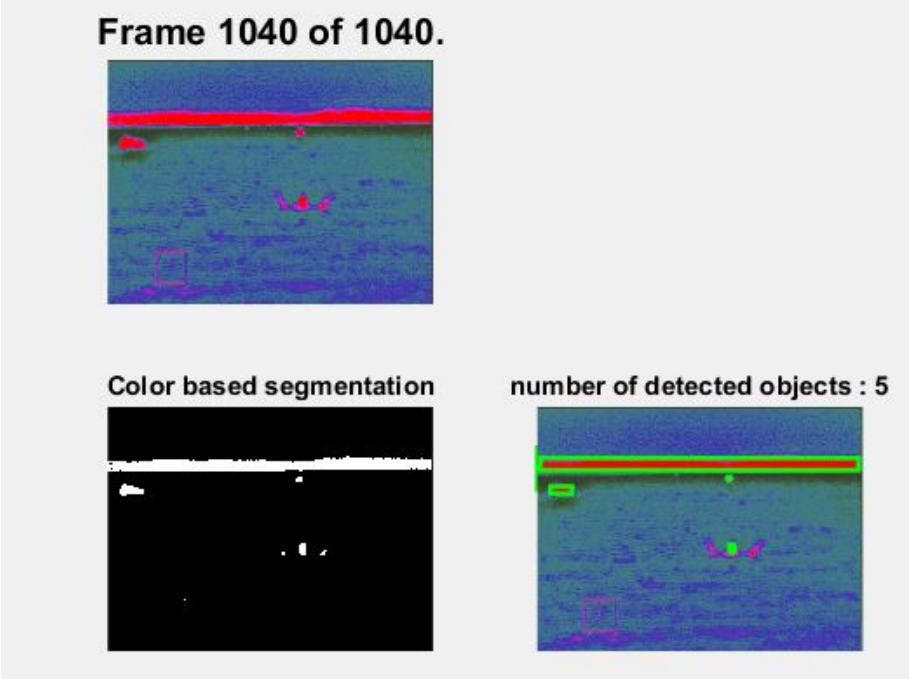


Figure 5.5: Detection in video

As we see in figure 5.6, we have processed 50 frames and number of True Positive frames are 46 and number of False Negative frames are 4.As we see in first frame number 1052 all the detection has been achieved successfully but in another frame which is number 1088 , there are 2 objects present in frame but only one has been detected so there is False Negative detection. Total CPU time to process it is 167389 sec and per frame average CPU time is 0334 sec.

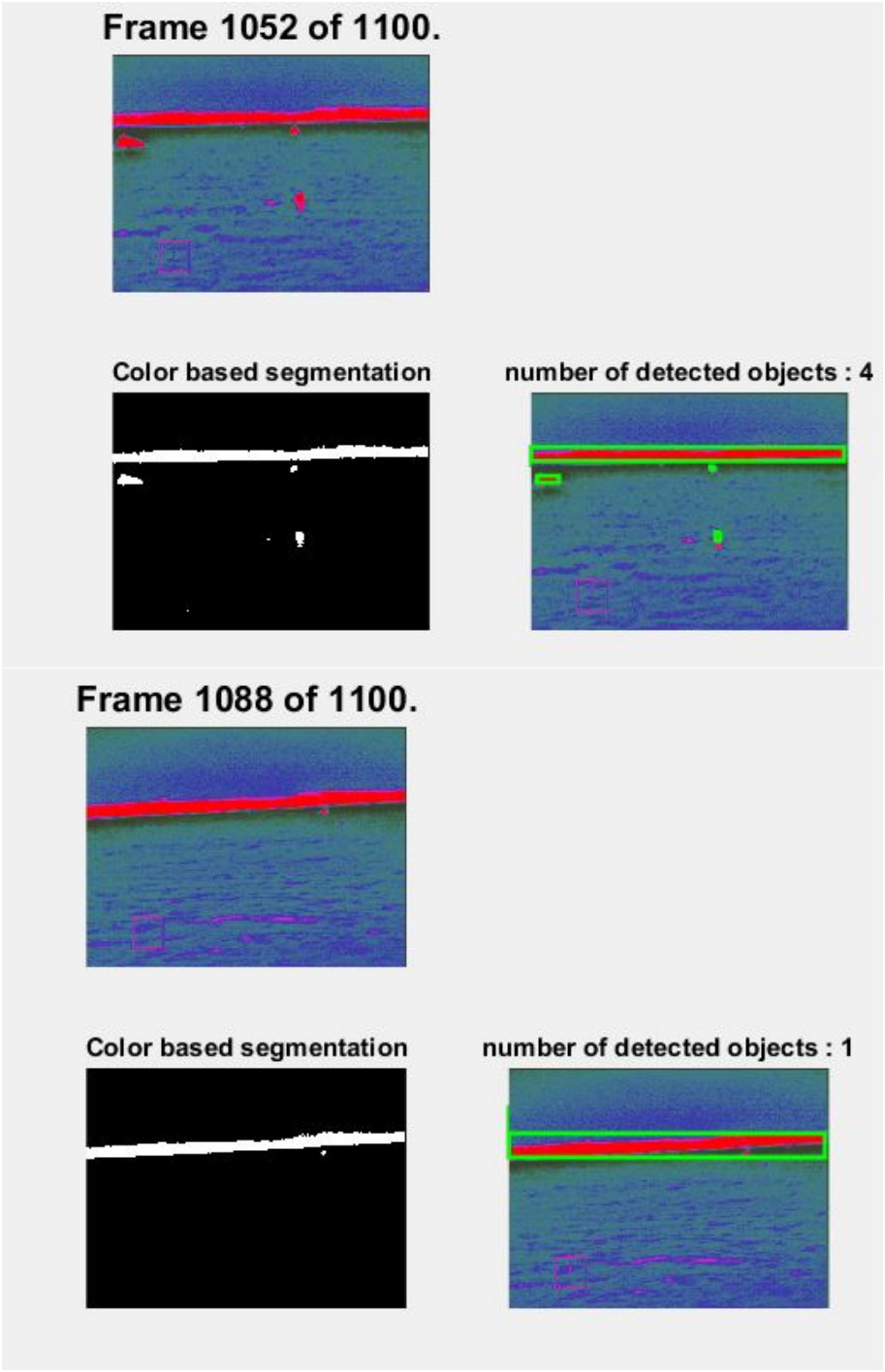


Figure 5.6: Detection in video

Table 5.1: Processing time

Type of input	No. of frames	Total CPU time	Average CPU time per frame
Image	1	0.421 sec	0.421 sec
Image	1	0.405 sec	0.405 sec
Image	1	0.37 sec	0.370 sec
Video	50	16.7389 sec	0.334 sec
Video	100	40.2483 sec	0.402 sec
Video	150	61.105 sec	0.407 sec

Table 5.2: Analysis

Type of input	No. of frames	TP frames	FP frames	FN frames
Image 2290	1	1	0	0
Image 1943	1	1	0	0
Image 3	1	0	1	0
Video	10	9	1	0
Video	20	16	2	2
Video	50	46	2	4

Chapter 6

Conclusion and future work

In this thesis we present the work towards object detection on water which can be used for search and rescue missions.

The ROAZ II autonomous surface vehicle has infrared and color camera which is used to collect raw data and after processing data we have detected objects in image or video. We can perform the algorithm on two type of images or video and that are color images or thermograhic images. If the video is very shaky then we can stabilize the video first and then we perform detection algorithm . After performing the stabilization we can see video is more stable and we can get better results in detection because we perform the task on water and on water surface because of waves it's difficult to get perfect detection. So to remove this problem stabilization has been used first. The algorithm is robust and very accurate in stabilization and detection. Stabilization removes the effect of jitter which is caused due to shaking of camera during video recording in sea or lakes. And detection removes unnecessary noise and from the results which we have obtained shown promising results and detect even very tiny object on water surface. Computation time is comparatively less as you can see in the above table 6.1. And algorithm is more precised as we can see from table 6.2.

In future we can perform detection in real time and object detection in world as it doesn't require stabilization. Because of stabilization the position of original objects may differ bit and due to that object avoidance will not be achieved successfully. And we

can achieve object tracking and object avoidance in world and in real time. Still we don't know the distance from ROAZ-II to detected object, so we can calculate that and it can be helpful for tracking and avoidance. And also we can perform this task in ROS because it is more robust to get higher detection speed than MatLab.

Chapter 7

Bibliography

[1] "ROAZ II Robot web page." [Online]. Available: http://lsa.isep.ipp.pt/pages/roaz$_r oaz$2.htm

[2] Michael T. Wolf, Christopher Assad, Yoshiaki Kuwata, Andrew Howard, Hrand Aghazarian, David Zhu, Thomas Lu, Ashitey Trebi-Ollennu, and Terry Huntsberger Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109. 360-Degree Visual Detection and Target Tracking on an Autonomous Surface Vehicle.

[3] Kalyanee G. Barve, S.S.Lokhande Department of Electronics Telecommunication, Sinhgad College of Engineering, Savitribai Phule Pune University, Pune, India. Obstacle Detection from Unmanned Surface Vehicle in International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering.

[4] Han Wang, Xiaozheng Mou, Wei Mou, Shenghai Yuan, Soner Ulun, Shuai Yang and Bok-Suk Shin School of Electrical and Electronic Engineering Nanyang Technological University Singapore 639798. Vision based Long Range Object Detection and Tracking for Unmanned Surface Vehicle

[5] Domenico Bloisi (a), Luca Iocchi (b), Michele Fiorini (c), Giovanni Graziano (d), (a)(b) Department of Computer and System Sciences Sapienza University of Rome, (c)(d) SELEX Sistemi Integrati S.p.A. - Roma. AUTOMATIC MARITIME SURVEILLANCE WITH VISUAL TARGET DETECTION

[6] Nuno Pires, Jonathan Guinet and Elodie Dusch Automatic Sea Vision 65 rue de la Garenne 92310 Sevres. ASV : An innovative automatic system for maritime surveillance.

[7] Alfredo Martins*, Andre Dias*, Jos e Almeida*, Hugo Ferreira*, Carlos Almeida*, Guilherme Amaral*, Diogo Machado*, Joao Sousa*, Pedro Pereira*, An ˜ ıbal Matos+, Vitor Lobo†, Eduardo Silva* +*INESC TEC Institute for Systems and Computer Engineering of Porto *ISEP - School of Engineering, Porto Polytechnic Institute, Porto, Portugal +Department of Electrical and Computer Engineering, Faculty of Engineering, University of Porto, Porto, Portugal,†CINAV, Portuguese Navy Research Center, Almada, Portugal. Field experiments for marine casualty detection with autonomous surface vehicles

[8] Alfredo Martins, Guilherme Amaral, Andre Dias, Carlos Almeida, Jose Almeida, Eduardo Silva. INESC TEC Robotics Unit, ISEP - School of Engineering Polytechnic Institute of Porto. TIGRE - An autonomous ground robot for outdoor exploration

[9] "Video Stabilization Using Point Feature Matching web page." [Online]. Available: http://www.mathworks.com/help/vision/examples/video-stabilization-using-point-feature-matching.html /

[10] "Video Stabilization Using Point Feature Matching in OpenCV web page." [Online]. Available: http://www.learnopencv.com/video-stabilization-using-point-feature-matching-in-opencv/

[11] Yadwinder Singh ,Lakhwinder Kaur . Punjabi University/Computer Engineering Department, Patiala, 147001, India. Obstacle Detection Techniques in Outdoor Envi-ronment: Process, Study and Analysis

[12] "HSL and HSV web page." [Online]. Available: http://en.wikipedia.org/wiki/HSL$_a nd_H SV/$

[13] "What is camera calibration web page." [Online]. Available: http://www.mathworks.com/help/vis calibration.html/

[14] "Calculate X, Y, Z Real World Coordinates from Image Coordinates using OpenCV web page." [Online]. Available: http://calculate-x-y-z-real-world-coordinates-

from-a-single-camera-using-opencv/

[14] Zhixiang Liua, Youmin Zhanga, Xiang Yua, Chi Yuan Department of Mechanical and Industrial Engineering, Concordia University, Montreal, Quebec H3G 1M8, Canada Unmanned surface vehicles: An overview of developments and challenges./

https://www.overleaf.com/project/5c8fc1907d18a5309e5fb7a8