AN EXPECTED ENTROPY REDUCTION (EER) APPROACH TO TARGET TRACKING IN MARITIME ENVIRONMENT FOR ON-BOARD CAMERA

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Xinyu Gao

August 2020

ABSTRACT

Maritime surveillance system installed on-board is crucial in protecting commercial vessels worldwide, which suffer potential losses due to the maritime piracy assault. On-board camera provides a way of monitoring suspicious activities at a low cost. However, the maritime environment poses challenges in detecting mobile targets because the water background is highly dynamic. This thesis addresses the problem of detecting and tracking mobile targets in the maritime environment utilizing an on-board camera. An approach based on optical flow is used to detect mobile targets presented in the camera scene. The camera measurement model and target kinematics model are proposed such that targets can be tracked utilizing the Bayesian filtering technique. A sensor planning strategy based on expected entropy reduction (EER) is developed to select optimal sensor field of view (FoV) locations such that the expected entropy reduction is maximized.

BIOGRAPHICAL SKETCH

Xinyu Gao is an M.S. student in the Laboratory for Intelligent Systems and Controls (LISC) at Cornell University. He received his B.S. degree in Mechanical Engineering from Purdue University. His research interests include robotics, computer vision, and probabilistic filtering.

This document is dedicated to my parents, Yaqing and Xiangyu for their love, support and encouragement all along the way.

ACKNOWLEDGEMENTS

I would like to express my thankfulness to Professor Silvia Ferrari, my advisor, for her constant motivation and support during my graduate study at Cornell University. Thanks to Professor Bharath Hariharan for being my defense committee and providing suggestions on my research and defense. I would like to thank Keith and Dr. Zhu for providing me valuable feedback on my research. Last but not least, I would like to thank all other members in LISC, Dr. Liu, Jake, Taylor, Julian, Yucheng, Shi, Hengye, Rui, Junyi, Jane, Dongheng, Yifeng, Shenghao, Zhihao for their support in the lab.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND BACKGROUND

Maritime piracy continues to threaten the sailors' safety and brings substantial economic losses to commercial vessels worldwide. It is crucial for people on-board to detect suspicious activities at sea in an early stage to reduce the risk of robbery [1]. Visual sensor surveillance system provides a way of detecting and tracking mobile targets in the sensor field of view at sea. Mobile sensors are utilized frequently in several applications such as surveillance system [2, 3] and environmental monitoring [4, 5]. Sensor planning considers the problem of developing a strategy that satisfies a sensing objective based on sensor measurements [6]. However, due to the dynamic nature of the water background caused by wakes, foams, etc, the maritime environment presents several challenges in mobile object detection and tracking. This thesis addresses the problem of detecting and tracking mobile objects in the maritime environment.

Maritime visual surveillance becomes a crucial topic in response to the rise of maritime piracy and possible collisions between vessels. Visual camera sensors installed on-board are considered for maritime surveillance purposes due to their affordable cost and efficiency in detecting smaller targets compared with radar sensors. The general framework of object detection applications consists of three major steps: horizon detection, background subtraction, and foreground segmentation [7]. However, typical approaches in detection and tracking utilizing background modeling and subtraction could perform poorly due to the dynamic background in the maritime environment [8]. Compared with the stationary background, water background poses challenges for several reasons. According to [9], background subtraction methods assume that the back-

ground is either spatially restricted or temporally restricted. In contrast, the water background caused by waves is changing continuously in both spatial and temporal dimensions. Secondly, dynamic background subtraction approaches indicate highly spatial and temporal corrections as foreground objects, which is not suitable for water background [10]. Thirdly, typical background subtraction methods show wakes and foams generated by vessels as foreground detection whereas in maritime environment those are background [9].

Several approaches have been developed to address the challenge presented in dynamic water background. Modified Boolean Map Saliency method with adaptive hysteresis thresholding introduced in [1] reduces the false detection generated by dynamic backgrounds such as waves and foams, but has problems detecting objects that are dark and far away. The foreground detection method utilizing the region-based Mixture of Gaussians modeling introduced by [11] reduces false positives presented in the dynamic scenes, which improved the typical Mixture of Gaussian modeling method. In [12], an improved optical flow and segmentation algorithm is developed based on the combination of color and motion information. Given the dynamic nature of the background, an object detection and segmentation algorithm is developed by [8] that utilizes a robust Kalman filter algorithm to model the dynamic textures presented in the background. This thesis addresses the mobile object detection problem under the dynamic maritime environment by using the Lucas-Kanade method from optical flow estimation [13].

After detecting the mobile object in the image frame from the video sequence, it is crucial to estimate the location and motion of the moving target in the world coordinates system based on those measurements. Tracking esti-

mates the state of moving targets based on measurements obtained by one or more sensors [14]. The Bayesian approach is studied to solve the tracking problem. The Bayesian approach takes the state as a random variable. The posterior probability density function (PDF) of state can be obtained from the state's prior PDF using Bayes' formula. Several typical tracking approaches have been developed such as Kalman filter [14] and particle filter [15]. Filtering refers to the problem of estimating the state of a dynamic system [14]. Estimation of the state could be obtained by filtering out the noisy measurement data. Kalman filter is a linear system estimator that requires both of its kinematics and measurement model to be linear. On the other hand, the Extended Kalman filter relaxes the linearity assumption such that it allows nonlinear kinematics and measurement model. Particle filter, one of the nonparametric Bayes filters, does not need a Gaussian representation of the states. Nonparametric filters could represent the posterior PDF by a set of random samples, which is capable of representing a broader state distribution compared with Gaussian based filters. However, those methods are more computationally expensive as the sample size increases. A review for visual tracking utilizing Kalman Filter is developed by [16]. A real-time video surveillance system developed by [17] that detects mobile objects based on a background model of water and tracks targets utilizing multiple hypothesis tracking by a set of Kalman filters. A target detecting and tracking method for a video surveillance system under the stationary background is developed by [18] where it utilizes extended Kalman filter algorithm for tracking purposes.

Given a region of interest, where the mobile target might be, the objective is to develop a sensor planning strategy such that the target can be tracked within the mobile sensor field of view (FoV) given imperfect sensor measurements [2].

Mobile sensors have the advantage of covering a larger area of region of interest (ROI) compared with static sensors, and their paths can be planned for possible target locations in the future [3]. The sensor FoV is a closed and bounded subset of ROI. The information about mobile targets within FoV provides a way of updating targets' state estimates by the Bayesian approach [19]. Tracking mobile targets with the sensor can be seen as a decision-making problem where the sensor needs to decide what to measure based on measurement information to optimize its tracking performance [20]. The information-driven approach provides a way of planning sensor decisions based on expected information value [2]. The information-theoretic functions are intended to measure the uncertainty presented in a random variable where in the case of tracking, the uncertainty of the target state is measured. The information-driven approach was introduced in [21] in solving sensor selection and object tracking problems where several measures of information utility were discussed such as entropy [22], Mahalanobis distance [23], and expected posterior distribution [24]. A greedy algorithm based on information value function introduced in [25] is used for sensor planning that allows sensing location to be selected based on maximizing the information value function given the measurement. The performance of several information-driven search strategies has been investigated using a target classification problem by [26], and the quadratic entropy-driven approach outperforms the others in terms of rates of correct-classification and the rate of false alarms. The performance of several information-theoretic functions is evaluated in solving the problem of sensor planning for tracking maneuvering targets, and the conditional mutual information-based objective function generates the most effective sensor planning approach [20]. A camera control approach is presented by [27] to follow unknown numbers of targets given limited

information by maximizing DP-GP information value function.

In this thesis, an optical flow estimation method is used to detect mobile targets in the maritime environment. The Bayesian approach is applied to track the mobile targets based on the estimated sensor measurement model and target kinematics model. A sensor planning strategy based on expected entropy reduction is presented to select sensor FoV location optimally so that the uncertainty of the target state could be minimized.

In Chapter 3, the optical flow method for detecting mobile targets presented in video sequences is discussed. As demonstrated in Chapter 4, a camera measurement model and the target kinematics model are proposed for the video surveillance system. A cost function is defined such that a set of optimal model parameters can be obtained given limited information from the video dataset. In Chapter 5, a Bayesian filtering algorithm based on the Extended Kalman filter is developed for tracking mobile targets. A sensor planning strategy based on expected entropy reduction (EER) is developed such that optimal sensor locations can be selected to minimize the target state uncertainty. As shown in Chapter 6, the proposed method allows the sensor to detect and track mobile targets in the maritime environment effectively.

CHAPTER 2

PROBLEM FORMULATION AND ASSUMPTIONS

This thesis considers the problem of detecting and tracking mobile targets in the maritime environment using an on-board pan and tilt camera. A sensor planning strategy based on expected entropy reduction is utilized to select optimal sensor locations.

The targets are moving in a workspace $\mathcal{W} \subset \mathbb{R}^2$. Let $\mathcal{F}_\mathcal{W}$ represents a fixed Cartesian frame in $\mathcal{W}$. Let $\mathcal{T}_i$, $i \in \{1, ..., n\}$ denotes ith target. The targets are modeled by Markov motion models, where the PDF of current target states at time step k can be calculated only from target states at the previous time step $(k-1)$. The target $\mathcal{T}_i$'s trajectory can be described as following by a time-invariant nonlinear ordinary differential equation:

$$\dot{\mathbf{x}}_i(t) = \mathbf{f}_i[\mathbf{x}_i(t)] \triangleq \mathbf{v}_i(t) \tag{2.1}$$

where $\mathbf{x}_i = \begin{bmatrix} x_i & y_i \end{bmatrix}^T \in \mathcal{W}$ and $\mathbf{v}_i = \begin{bmatrix} \dot{x}_i & \dot{y}_i \end{bmatrix}^T \in \mathbb{R}^2$ denote the position and velocity of target i in inertial frame $\mathcal{W}$.

The target's state is defined as $\mathbf{x}_T = \begin{bmatrix} x & y & \dot{x} & \dot{y} \end{bmatrix}^T$. Assuming each target preserves constant velocity, the kinematics model of target i can be described as following equation where the target state transition is established between kth step and $(k-1)$th step.

$$\mathbf{x}_T(k) = \mathbf{F}\mathbf{x}_T(k-1) \tag{2.2}$$

where,

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.3}$$

Δt denotes the constant time interval.

A pan and tilt camera installed on a vessel at sea is assumed to be at a fixed location with respect to inertial frame $\mathcal{W}$. The sensor field of view (FoV) is a closed and bounded subset of the region of interest (ROI) where its size is largely reduced compared with the camera scene [19], denoted as $\mathcal{S} \subset \mathbb{F}_p$, where the $\mathcal{W}$ is projected onto a virtual image plane $\mathbb{F}_p$ that can obtain the measurement of mobile targets. According to the pinhole camera model [28], camera images are projected onto the virtual image plane $\mathbb{F}_p$, where the distance between the image plane $\mathbb{F}_p$ and pinhole is denoted as λ (the focal length). Let $\mathcal{F}_A$ denotes the Cartesian frame of the camera sensor, and O_A, the pinhole location, indicates the origin of this frame. Camera measurements of mobile targets can be obtained within ROI by optical flow estimation at each time step k given a video sequence. A measurement model based on pan-tilt camera model can be described as following,

$$\mathbf{y}(k) = \begin{cases} \mathbf{h}(\mathbf{x}(k)) + \mathbf{w}_k & \text{if} \quad \mathbf{p}_T(k) \in \mathcal{S}(k) \\ \emptyset & \text{if} \quad \mathbf{p}_T(k) \notin \mathcal{S}(k) \end{cases} \tag{2.4}$$

where $\mathbf{h}(\cdot)$ denotes the measurement function and $\mathbf{w}_k \in \mathbb{R}^2$ indicates the additive Gaussian noise with zero mean and $\mathbf{p}_T = \begin{bmatrix} p_x & p_y \end{bmatrix}^T$ denotes the location of target in image frame $\mathbb{F}_p$.

In order to track mobile targets in the sensor FoV, the Bayesian filtering technique based on the extended Kalman filter is utilized. Unknown camera measurement model parameters and target kinematics model parameters are estimated from the ground truth dataset. Gaussian distribution is assumed for the target states. Current target states can be estimated in two steps by the target states at the previous time step $(k-1)$. First, the predicted distribution of the target state is calculated based on state posterior distribution at the previous time step $(k-1)$ given the target dynamics. Secondly, the state posterior distribution at the current time step k is updated by the measurements observed from the camera sensor.

To select sensor FoV (largely reduced) location optimally inside the camera scene to obtain measurements that maximize the target information based on prior target distribution, a sensor planning strategy utilizing expected entropy reduction (EER) is proposed. Given a camera scene, the objective is to place the simulated FoV $\mathcal{S}$ to an optimal location where it can obtain the most information of the target. The expected entropy reduction [2] can be represented as following over the time interval $[t_i, t_f]$. The problem is formulated to maximize the cost function J.

$$J = \sum_{k=i}^{f} \int_{\mathcal{Z}_k} p(\mathbf{z}(t_k) \mid \mathbf{M}^{k-1}) R[\mathbf{z}(t_k)] d\mathbf{z}(t_k) \tag{2.5}$$

where $\mathcal{Z}_k$ indicates the range of measurement $\mathbf{z}(t_k)$. The conditional measurement PDF is denoted as $p(\mathbf{z}(t_k) \mid \mathbf{M}^{k-1})$ and the set of available measurements at time t_k is $\mathbf{M}^{k-1}$. The reward function denoted as $R[\mathbf{z}(t_k)]$ measures the reduction in differential entropy.

The primary dataset investigated in this thesis is from PEST 2016 dataset

[29], where a video sequence is generated from the visual sensor in the maritime environment. A specific video frame is shown in Figure 2.1, and a simulated FoV is presented in the red bounding box.

Figure 2.1: An example of a camera scene in the video sequence where the red bounding box represents the simulated sensor's FoV.

CHAPTER 3

TARGET DETECTION VIA OPTICAL FLOW ESTIMATION

Marine piracy brings considerable costs to the economy and humans on-board over the years, making the on-board sensor surveillance critical in detecting suspicious activities. This chapter considers the problem of detecting mobile targets in the maritime environment. One of PETS 2016 datasets [29] provided camera video recording that serves the purpose of detecting and tracking mobile targets at sea. The optical flow estimation method is used for detecting moving targets. A bilateral filter is applied to smooth the image and reduce the noise due to the dynamic background caused by waves.

3.1 Description of Video Data Sets

In this video dataset, two speedboats (moving targets) appear from the top-right corner together. They separate in the middle of the video with one leading the other. Finally, they disappear from the left side.

Figure 3.1: The snapshot of the video where two speedboats are in the region of interest.

3.2 Optical Flow Estimation for Boat Detection

Motion estimation, known as optical flow, can be performed given sequences of time-ordered images where the velocity vector for each pixel is calculated from one frame to another frame. In order to detect the speedboats in each image frame, gradient based Lucas-Kanade method [13] from optical flow estimation is used. The Lucas-Kanade method performs fast calculation with precise time derivatives compared with other methods [30]. The Lucas-Kanade method assumes that the displacement of the image between two nearby frames is small and it solves the basic optical flow equations for all the pixels. The local image velocity (V_x, V_y) should satisfy the following equation [13].

$$\begin{aligned}
\mathbf{I}_x(q_1)V_x + \mathbf{I}_y(q_1)V_y &= -\mathbf{I}_t(q_1) \\
\mathbf{I}_x(q_2)V_x + \mathbf{I}_y(q_2)V_y &= -\mathbf{I}_t(q_2) \\
&\vdots \\
\mathbf{I}_x(q_n)V_x + \mathbf{I}_y(q_n)V_y &= -\mathbf{I}_t(q_n)
\end{aligned} \tag{3.1}$$

where q_1, q_2, ..., q_n are pixels inside the neighborhood, and $\mathbf{I}_x(q_n)$, $\mathbf{I}_y(q_n)$, $\mathbf{I}_t(q_n)$ indicate the partial derivatives of image $\mathbf{I}$ with respect to position x, y and time t at point q_n.

These equations could be written in terms of matrix form $\mathbf{Av} = \mathbf{b}$. The Lucas-Kanade method then solve these equations by the least squares criterion.

$$\begin{aligned}
\mathbf{A}^T\mathbf{Av} &= \mathbf{A}^T\mathbf{b} \\
\mathbf{v} &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}
\end{aligned} \tag{3.2}$$

$$\begin{bmatrix} \mathbf{V}_x \\ \mathbf{V}_y \end{bmatrix} = \begin{bmatrix} \Sigma_i I_x(q_i)^2 & \Sigma_i I_x(q_i) I_y(q_i) \\ \Sigma_i I_y(q_i) I_x(q_i) & \Sigma_i I_y(q_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\Sigma_i I_x(q_i) I_t(q_i) \\ -\Sigma_i I_y(q_i) I_t(q_i) \end{bmatrix} \tag{3.3}$$

For the boat video dataset described above, the background noise generated by sea foams and waves is relatively large so that the image needed to be filtered to reduce the noise. Figure 3.2 shows the velocity field of the unfiltered image.

Figure 3.2: The snapshots of the video at two different moments, where velocity field vectors are extracted shown in red arrows.

Water background, one of the dynamic backgrounds, provides a complicated environment for foreground detection where the sea foams and wakes that have relatively large motions should be considered as background. In this specified video dataset, two types of background noises (white foams generated by speedboats and sea foams that moving in a constant direction) should be considered. The sea foams have a consistent moving direction, which is moving away vertically in the video. Therefore, given the flow orientation calculated by the Lucas-Kanade method at each pixel, the effect of white foams could be largely eliminated. The white foams generated by the speedboats should be taken as the background, although given its relatively large motion compared with sea foams. A bilateral filter is used to reduce the noise and to smooth the image before calculating the optical flow. Compared with domain filters such as

Gaussian low-pass filter that averages the pixel value within a neighborhood at every pixel so as to average the noise presented in image away, the bilateral filter is a nonlinear noise reduction filter that preserves edge in a given image [31]. The intensity value of each pixel is replaced by the weighted average intensity value of the nearby pixels while preserving edges by considering the difference in color between the current pixel and neighbor pixels.

Once the bilateral filter is applied to reduce the noise and velocity vector field is calculated by the Lucas-Kanade method, foreground segmentation can be made by assigning rectangular bounding boxes to different groups of velocity vector which indicate the moving targets given an image frame. The center of the bounding box is taken as the location of moving targets.

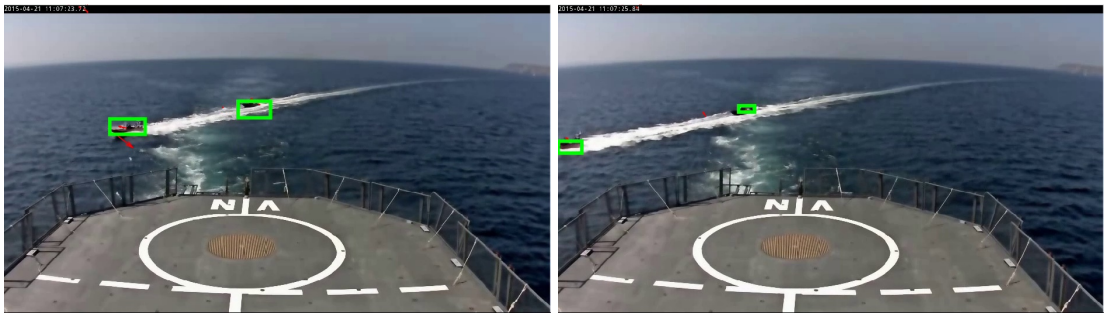


Figure 3.3: The snapshots of the video at two different moments, where two speedboats were detected. Green rectangles indicate bounding boxes.

CHAPTER 4

PARAMETER ESTIMATION FOR CAMERA MEASUREMENT AND TARGET KINEMATICS MODEL

In this chapter, the camera measurement model and the target kinematic model are defined so that the target can be tracked using the Bayesian filtering techniques. The camera is located on a moving vessel given the video dataset description given in the previous chapter. A pan-tilt camera model is assumed. Given the dynamics of the moving targets from the dataset, a unicycle kinematics model is assumed. Given that the camera model parameters and target kinematic model parameters are unknown in this scenario, a cost function is defined such that a set of optimized parameters can be found based on the ground truth measurement data.

4.1 On-board Camera Sensor

2016 PETS (Performance Evaluation of Tracking and Surveillance) datasets consider the problem of protecting mobile assets via the on-board multi-sensors surveillance approach. Two datasets are provided, land-based dataset and maritime dataset. This thesis adapts the maritime dataset, where the on-board camera shows the moving boat targets that need to be detected and tracked. An AXIS P1427-E Network camera was installed at the stern of the recording vessel [29]. Figure 4.1 shows the sensor locations on the recording vessel.

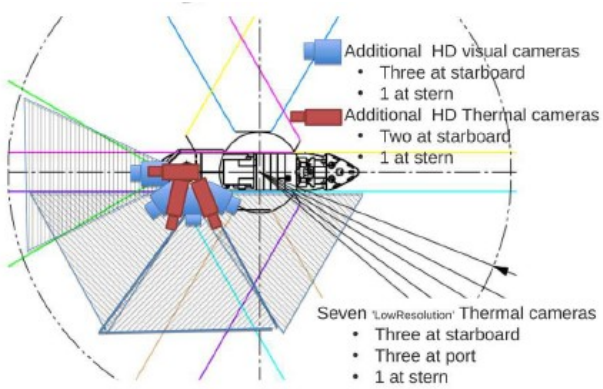


Figure 4.1: PESTS2016 dataset sensor location. Source: Adapted from [29].

4.2 Image Distortion Correction on Video Dataset

A fish-eye lens is equipped such that the field of view is wider compared with the standard lens. However, the image may be distorted due to the effect of deviation from the rectilinear projection. In the case of this maritime video dataset, barrel distortion occurred in the images. As shown in Figure 4.2, the sea level in the field of view is curved, which should be a straight line if there is no image distortion. An iterative optimization algorithm for lens distortion correction is proposed by [32] and used for finding distortion parameters.

Figure 4.2: Image distortion presented in the camera field of view.

The general equation for lens distortion model is given by following equation,

$$\begin{bmatrix} \hat{x} - x_c \\ \hat{y} - y_c \end{bmatrix} = L(r) \begin{bmatrix} x - x_c \\ y - y_c \end{bmatrix} \tag{4.1}$$

where (x_c, y_c) represents the distortion center. (x, y) indicates pixel coordinates in the original image where the point is distorted. $(\hat{x}, \hat{y})$ denotes the undistorted pixel coordinates. $L(r)$ defines the type of distortion model function and $r = \|(x, y) - (x_c, y_c)\|$ [33]. The polynomial model and the division model are two commonly used radial distortion models [32]. The polynomial model is used here to correct the distorted image and is shown in the following equation,

$$L(r) = 1 + k_1 r^2 + k_2 r^4 \tag{4.2}$$

where k_1 and k_2 are the distortion parameters.

Additional parameters p_1 and p_2 are introduced instead such that k_1 and k_2 can be normalized [32], where p_1 represents percentage of correction from the furthest point in the image to the distortion center and p_2 indicates same percentage of correction but uses half the distance between furthest point and distortion center. The equations that describe the relationship between two sets of parameters are shown below,

$$\begin{aligned} (1 + p_1) r_1 &= r_1 (1 + k_1 {r_1}^2 + k_2 {r_1}^4) \\ (1 + p_2) r_2 &= r_2 (1 + k_1 {r_2}^2 + k_2 {r_2}^4) \end{aligned} \tag{4.3}$$

Where r_1 denotes the distance between the center of the distorted image, and furthest point in the image and r_2 indicates the distance that is half of r_1. The

Equation 4.3 can be simplified as,

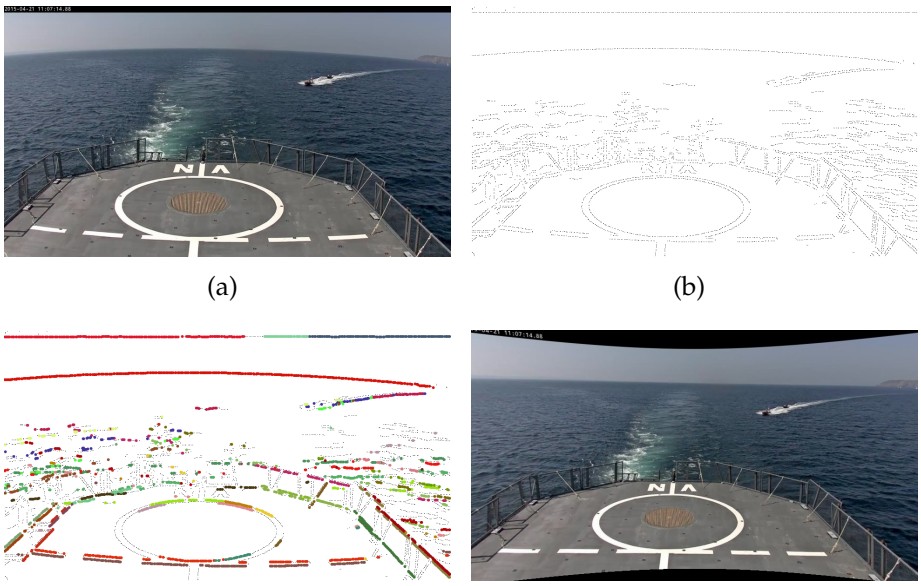
$$p_1 = k_1 4{r_2}^2 + k_2 16{r_2}^4$$
$$p_2 = k_1 {r_2}^2 + k_2 {r_2}^4 \tag{4.4}$$

k_1 and k_2 can be further expressed in terms of p_1, p_2 and r_2,

$$k_1 = \frac{p_1 - 16p_2}{-12{r_2}^2}, k_2 = \frac{4p_2 - p_1}{-12{r_2}^4} \tag{4.5}$$

This automatic radial distortion correction algorithm is achieved in four steps [32]. First, the canny edge detector is used to extract edge points in the image so that the distorted lines can be located. Second, a modified Hough transform method is proposed to detect the straight line in the original image. Third, an iterative optimization algorithm is performed to optimize the lens distortion model parameters and to detect points that were not considered before using the Hough transform. Forth, given the optimized lens distortion model parameters, the corrected image is calculated based on the inverse of the distortion model.

Given an uncorrected boat image, a set of optimized lens distortion model parameters are calculated using the automatic distortion correction algorithm described in [32], and the corrected image is shown in Figure 4.3. By applying lens distortion model parameters to each image frame, a corrected image distortion target trajectory is obtained based on the ground truth boat detection dataset, which can be shown in Figure 4.4.



(a) (b) (c) (d)

Figure 4.3: (a) presents the distorted boat dataset image; (b) shows the canny edge detector result; (c) demonstrates lines detection using Hough transform; and, (d) indicates the corrected image.

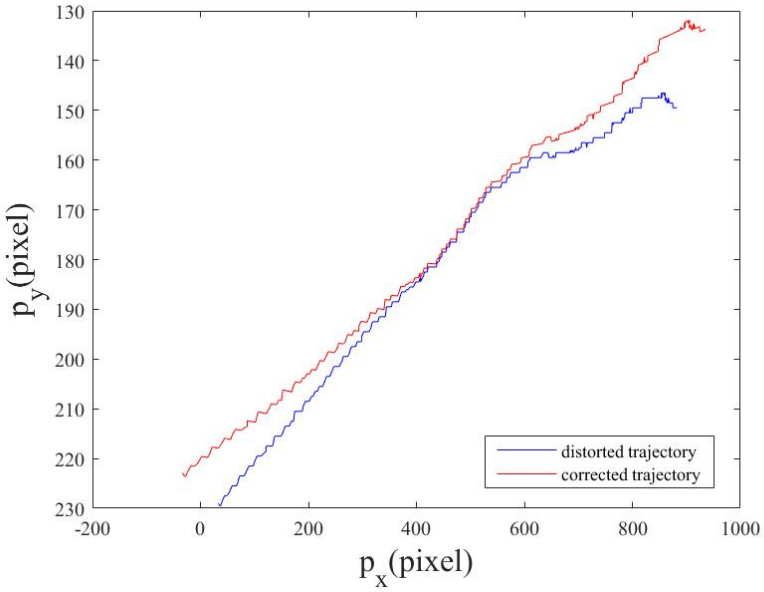


Figure 4.4: Blue line indicates the original boat trajectory in the image frame. Red line demonstrates the corrected distortion trajectory in the image frame.

4.3 Sensor Measurement Model

Sensor modeling is an essential step in detecting and tracking mobile targets [2]. In order to model the video surveillance system in the maritime environment as described in the previous section, a pan-tilt camera model is assumed [34]. The camera sensor is assumed to be at a fixed position with respect to the world inertial frame $\mathcal{W}$. The field of view of the camera denoted as $\mathcal{S}(k) \subset \mathbb{F}_p$ is the bounded region that allows the camera to take measurements at each time step k. A pinhole camera model allows points presented in 3D inertial frame to be transformed into a virtual image frame, which is illustrated in Figure 4.5. The focal length λ is the distance between the virtual image plane and the optical center of the camera. Let $\mathbf{x}_0 = \begin{bmatrix} x_0 & y_0 & z_0 \end{bmatrix}^T$ denotes the pinhole position [34] with respect to the inertial frame and $\mathbf{x}_i \in \mathcal{W}$ denotes the location of the moving target i. The pan and tilt angle of the camera are denoted as yaw (ψ) and roll (ϕ) angles according to Euler angle representation which can be illustrated in Figure 4.6. The position of target i with respect to camera coordinates is shown below,

$$\mathbf{q}_i = \mathbf{R}_\phi \mathbf{R}_\psi (\begin{bmatrix} \mathbf{x}_i^T & 0 \end{bmatrix}^T - \mathbf{x}_0) = \begin{bmatrix} q_x & q_y & q_z \end{bmatrix}^T, \tag{4.6}$$

where $\mathbf{R}_\psi$ and $\mathbf{R}_\phi$ are transformation matrices corresponding to pan and tilt angle that are defined by,

$$\mathbf{R}_\psi \triangleq \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.7}$$

and

$$\mathbf{R}_\phi \triangleq \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix} \tag{4.8}$$

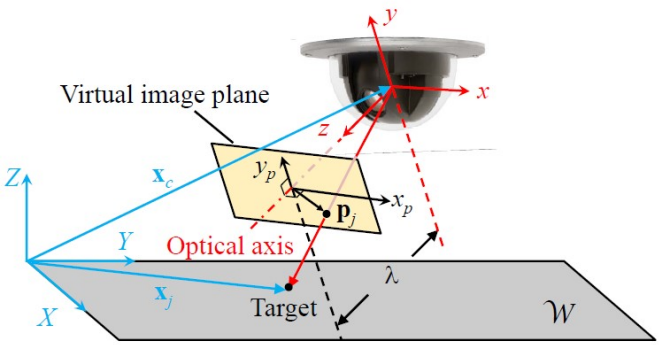


Figure 4.5: Pinhole camera model. Source: taken from [34] with permission.

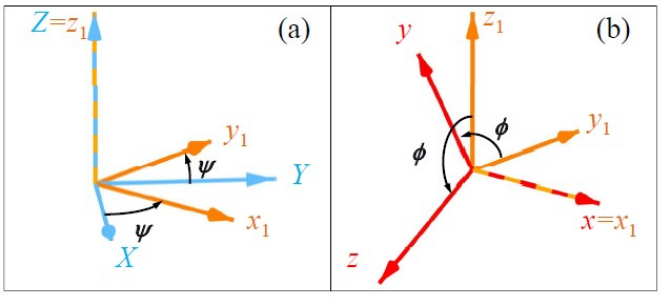


Figure 4.6: Rotation matrix. Source: taken from [34] with permission.

Next the target position $\mathbf{q}_i$ with respect to camera fixed frame can be projected onto virtual image plane as shown below,

$$\mathbf{p}_i = \begin{bmatrix} p_x & p_y \end{bmatrix}^T = \lambda \begin{bmatrix} \frac{q_x}{q_z} & \frac{q_y}{q_z} \end{bmatrix}^T \tag{4.9}$$

where p_x and p_y are pixel coordinates in image plane.

The origin of the virtual image plane is at the intersection between the camera z axis and image plane, where the digital image has the origin at lower left corner of the image. Therefore, there should be a translation vector $\begin{bmatrix} c_x & c_y \end{bmatrix}^T$ to encounter the offset [35]. The mapping becomes,

$$\mathbf{p}_i = \begin{bmatrix} p_x & p_y \end{bmatrix}^T = \begin{bmatrix} \lambda \frac{q_x}{q_z} + c_x & \lambda \frac{q_y}{q_z} + c_y \end{bmatrix}^T \tag{4.10}$$

The points in the digital image have the unit of pixel, while points described by the transformation above have the unit of physical measurement (e.g. such as centimeters). Two parameters k and l are introduced to count the unit change between two axes of image plane [35]. The mapping is adjusted to be:

$$\mathbf{p}_i = \begin{bmatrix} p_x & p_y \end{bmatrix}^T = \begin{bmatrix} \lambda k \frac{q_x}{q_z} + c_x & \lambda l \frac{q_y}{q_z} + c_y \end{bmatrix}^T = \begin{bmatrix} \alpha \frac{q_x}{q_z} + c_x & \beta \frac{q_y}{q_z} + c_y \end{bmatrix}^T \tag{4.11}$$

This coordinate transformation which allows a target location $\mathbf{x}_i$ in world inertial frame to be projected into camera virtual image plane $\mathbb{F}_p$ defines the measurement function: $\mathbf{h}(\cdot)$ such that it measures the target location in the image virtual frame. The objective is to learn the camera model and boat kinematics parameters from the observation data (ground truth boat trajectory).

$$\mathbf{h}(\mathbf{x}_i(k)) = \mathbf{p}_i(k) \tag{4.12}$$

4.4 Target Kinematics Model

Targets such as boats usually preserve a constant heading and velocity motion for an extended period of time in the maritime environment. The primary dataset investigated by this Thesis obtains a similar target motion. Therefore, it is appropriate to assume a unicycle kinematics model with constant velocity and heading for the targets of interests. A configuration is denoted by $\mathbf{x} = \begin{bmatrix} x & y \end{bmatrix}^T$. Let $\mathbf{x}_{init} = \begin{bmatrix} x_{init} & y_{init} \end{bmatrix}^T$ denotes the prior position of the target. The motion of the targets could be described by following set of equations.

$$\begin{aligned} x(k) &= x(k-1) + v_c \cos\theta_c \cdot \Delta t \\ y(k) &= y(k-1) + v_c \sin\theta_c \cdot \Delta t \end{aligned} \tag{4.13}$$

where $v_c \in (0, 12.85)[m/s]$ is the constant velocity of the target [29] and $\theta_c \in (0, 2\pi)$ is the constant heading angle of the target. We assume a constant time interval Δt.

4.5 Camera and Boat Kinematics Parameters Estimation

In order to optimize camera and boat kinematic model parameters for future boat location estimation, a cost function is defined as follows. The error between the boat trajectory obtained from the ground truth dataset and estimated trajectory by the measurement function needs to be minimized.

$$J_p = \min \sum_{k=1}^{n} \|\mathbf{h}(\mathbf{x}_i(k)) - \mathbf{z}_{gt}(k)\|^2 \tag{4.14}$$

where $\mathbf{h}(\cdot)$ is the measurement function defined in Equation 4.12, n represents number of observations, $\mathbf{x}_i(k)$ represents the position of the target i in inertial frame $\mathcal{W}$ and $\mathbf{z}_{gt}(k)$ indicates the pixel coordinates of the boat location in kth camera frame from ground truth dataset.

CHAPTER 5

TARGET TRACKING VIA EXPECTED ENTROPY REDUCTION

In this chapter, the Bayesian filtering techniques are studied in order to track the mobile targets presented in the maritime environment given the sensor measurement model and target kinematics model defined in the previous chapter. Once the conditional PDF is derived by one of the recursive Bayesian state estimators such as Kalman filter or particle filter, an information-driven sensor planning strategy, expected entropy reduction (EER), is applied such that the optimal sensor location is obtained to minimize the uncertainty of target states at each time step [2].

5.1 State Estimation Utilizing Bayesian Filtering Techniques

To predict the state of targets in the maritime environment, the Bayesian filtering techniques are applied. The Extended Kalman filter relaxes the linearity assumption of the Kalman filter that it allows the system to be nonlinear. In the boat tracking problem, the camera measurement function is nonlinear so that an extended Kalman filter is suitable for the tracking purpose.

5.1.1 Kalman filter and Extend Kalman filter

A brief review of the Kalman filter and Extended Kalman filter is presented here. For Kalman filter, the dynamic and measurement model are assumed to

be linear Gaussian which is given by the following equations,

$$\begin{aligned} \mathbf{x}_k &= \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \boldsymbol{\nu}_{k-1} \\ \mathbf{y}_k &= \mathbf{H}_k\mathbf{x}_k + \mathbf{w}_k \end{aligned} \tag{5.1}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ represents the state vector, $\mathbf{y}_k \in \mathbb{R}^m$ denotes the measurement vector. $\boldsymbol{\nu}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q_{k-1}})$ denotes the Gaussian process noise and $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R_k})$ represents the Gaussian measurement noise. The matrix $\mathbf{F}_{k-1}$ represents the state transition matrix of the dynamic model and matrix $\mathbf{H}_k$ is the measurement model matrix. There are two crucial steps in applying the Kalman filter, namely the prediction step and the update step. In the prediction step, the predicted distribution of state $\mathbf{x}_k$ is calculated by integrating the posterior distribution at state $\mathbf{x}_{k-1}$ given the dynamic model via Chapman-Kolmogorov equation [36].

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})d\mathbf{x}_{k-1} \tag{5.2}$$

where $\mathbf{y}_{1:k-1}$ represents all prior measurements up to time step $k-1$. In the update step, the posterior distribution of state $\mathbf{x}_k$ given the measurement $\mathbf{y}_k$ at time step k can be computed using Bayes' rule [36].

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{\int p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})d\mathbf{x}_k} \tag{5.3}$$

Since the measurement function of the camera sensor model defined in the previous chapter is nonlinear, we need to consider using extended Kalman filter to do the state estimation, where state transition function and the measurement function are represented by nonlinear functions denoted as $\mathbf{f}$ and $\mathbf{h}$.

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \boldsymbol{\nu}_{k-1} \tag{5.4}$$

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{w}_k \tag{5.5}$$

Due to nonlinearity of the function **f** and **h**, the extended Kalman filter calculates a Gaussian approximation to the true filtering distribution [37]. The extended Kalman filter linearized the nonlinear functions **f** and **h** by means of (first-order) Taylor expansion. Therefore the functions **f** and **h** can be described as following,

$$\mathbf{f}(\mathbf{x}_{k-1}) \approx \mathbf{f}(\mathbf{m}_{k-1}) + \mathbf{F_k}(\mathbf{x}_{k-1} - \mathbf{m}_{k-1}) \tag{5.6}$$

$$\mathbf{h}(\mathbf{x}_k) \approx \mathbf{h}(\mathbf{m}_k^-) + \mathbf{H_k}(\mathbf{x}_k - \mathbf{m}_k^-) \tag{5.7}$$

where $\mathbf{m}_{k-1}$ denotes the mean of posterior distribution and $\mathbf{m}_k^-$ indicates the prior distribution. State transition and measurement matrices, $\mathbf{G_k}$ and $\mathbf{F_k}$, are defined by following Jacobians,

$$\mathbf{F}_k = \left.\frac{\partial \mathbf{f}(\mathbf{x}_{k-1})}{\partial \mathbf{x}_{k-1}}\right|_{\mathbf{x}_{k-1}=\mathbf{m}_{k-1}} \tag{5.8}$$

$$\mathbf{H}_k = \left.\frac{\partial \mathbf{h}(\mathbf{x}_k)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k=\mathbf{m}_k^-} \tag{5.9}$$

The extended Kalman filter algorithm is shown below [36].
The prediction step:

$$\begin{aligned} \mathbf{m}_k^- &= \mathbf{f}(\mathbf{m}_{k-1}), \\ \mathbf{P}_k^- &= \mathbf{F}_k(\mathbf{m}_{k-1})\mathbf{P}_{k-1}\mathbf{F}_k^T(\mathbf{m}_{k-1}) + \mathbf{Q}_{k-1} \end{aligned} \tag{5.10}$$

where $\mathbf{m}_k^-$ is the mean of prior state estimate and $\mathbf{P}_k^-$ is prior covariance matrix.
The update step:

$$\begin{aligned} \mathbf{S}_k &= \mathbf{H}_k(m_k^-)\mathbf{P}_k^-\mathbf{H}_k^T(\mathbf{m}_k^-) + \mathbf{R}_k, \\ \mathbf{K}_k &= \mathbf{P}_k^-\mathbf{H}_k^T(\mathbf{m}_k^-)\mathbf{S}_k^{-1}, \\ \mathbf{m}_k^+ &= \mathbf{m}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{h}(\mathbf{m}_k^-)), \\ \mathbf{P}_k^+ &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^- \end{aligned} \tag{5.11}$$

where $\mathbf{m}_k^+$ is updated state estimate and $\mathbf{P}_k^+$ is updated covariance matrix.

5.1.2 Extended Kalman filter Implementation

Given the optimized camera sensor model parameters and target kinematics model parameter found in the previous chapter, tracking through extended Kalman filter can be formulated as follows. Let $\mathbf{x}_T = \begin{bmatrix} x & y & \dot{x} & \dot{y} \end{bmatrix}^T$ denotes the dynamic state of the boat, where $\mathbf{x} = \begin{bmatrix} x & y \end{bmatrix}^T \in \mathcal{W}$ is the target position in world inertial frame. The boat dynamic equation can be expressed in state-space form in the following,

$$\mathbf{x}_T(k) = \mathbf{F}\mathbf{x}_T(k-1) + \boldsymbol{\nu}(k-1) \tag{5.12}$$

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.13}$$

where $\boldsymbol{\nu}(k-1) \sim \mathcal{N}(\mathbf{0}, Q)$ and $\mathbf{F}$ denotes the state transition matrix assuming a constant time interval Δt.

Measurement function for the camera model is nonlinear that was defined in Equation 4.12 and is shown explicitly in below. $\mathbf{y}(k) = \mathbf{h}(\mathbf{x}(k)) + \mathbf{w}(k)$, where $\mathbf{w}(k) \sim \mathcal{N}(\mathbf{0}, \mathcal{R})$.

$$\mathbf{h}(\mathbf{x}(k)) = \begin{bmatrix} \alpha \frac{\cos(\phi)\cdot x + \sin(\psi)\cdot y}{\sin(\phi)\sin(\psi)\cdot x - \sin(\phi)\cos(\psi)\cdot y - \sin(\phi)\cdot z_0} + c_x \\ \beta \frac{-\cos(\phi)\sin(\psi)\cdot x + \cos(\phi)\cos(\psi)\cdot y - \sin(\phi)\cdot z_0}{\sin(\phi)\sin(\psi)\cdot x - \sin(\phi)\cos(\psi)\cdot y - \sin(\phi)\cdot z_0} + c_y \end{bmatrix} \tag{5.14}$$

Because of the non-linearity of the measurement function, prior state estimate is calculated via Taylor Expansion. The Jacobian matrices for measurement and state transition function $\mathbf{H}(k)$ and $\mathbf{F}(k)$ are shown below.

$$\mathbf{F}(k) = \frac{\partial \mathbf{f}(\mathbf{x}_T(k-1))}{\partial \mathbf{x}_T(k-1)} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.15}$$

$$\mathbf{H}(k) = \frac{\partial \mathbf{h}(x,y)}{\partial \mathbf{x}_T(k)} = \begin{bmatrix} \frac{\partial h_1(x,y)}{\partial x} & \frac{\partial h_1(x,y)}{\partial y} & \frac{\partial h_1(x,y)}{\partial \dot{x}} & \frac{\partial h_1(x,y)}{\partial \dot{y}} \\ \frac{\partial h_2(x,y)}{\partial x} & \frac{\partial h_2(x,y)}{\partial y} & \frac{\partial h_2(x,y)}{\partial \dot{x}} & \frac{\partial h_2(x,y)}{\partial \dot{y}} \end{bmatrix} \tag{5.16}$$

where $h_1(\cdot)$ represents the first term and $h_2(\cdot)$ represents the second term in the measurement function $\mathbf{h}(\mathbf{x}(k))$.

5.2 Optimal Sensor Placement via Expected Entropy Reduction

Given imperfect sensor measurements, sensor planning is crucial in placing sensor locations optimally to reduce the state's uncertainty. Once the conditional PDF of the target state is updated by the Bayesian filtering technique described above, the Expected Entropy Reduction method is used to select optimal sensor locations given the prior state estimates at each time step so that the uncertainty of the state can be minimized [2]. The posterior distribution is not available before obtaining the sensor measurements and thus the actual entropy reduction can not be calculated [6]. Therefore, the expected information value is calculated for the sensor to decide which measurement to take in the future [20]. Information-theoretic functions are used to quantify the information value of sensor measurement. Entropy is a common way of measuring information value [26]. Shannon entropy is defined as the following,

$$H(\mathrm{X}) = -\int_{\mathcal{X}} p(\mathrm{X}) \log_2 p(\mathrm{X}) d\mathrm{X} \tag{5.17}$$

where X is a continuous random variable with range $\mathcal{X} \subset \mathbb{R}$ and the probability density function is denoted as $p(\mathrm{X})$.

From state estimation point of view, the uncertainty in target state $\mathbf{x}(t_k)$, based on prior measurements $\mathbf{M}^{k-1}$, should be decreased by the measurement $\mathbf{z}(t_k)$ [2]. The conditional entropy at time t_k can be represented as following,

$$H(\mathbf{x}(t_k) \mid \mathbf{M}^{k-1}) = -\int_{\mathcal{X}} p(\mathbf{x}(t_k) \mid \mathbf{M}^{k-1}) \log_2 p(\mathbf{x}(t_k) \mid \mathbf{M}^{k-1}) d\mathbf{x}(t_k) \tag{5.18}$$

A reward function, which is also called conditional mutual information, is defined such that the reduction of entropy can be measured given measurement $\mathbf{z}(t_k)$.

$$R[\mathbf{z}(t_k)] = H(\mathbf{x}(t_k) \mid \mathbf{M}^{k-1}) - H(\mathbf{x}(t_k) \mid \mathbf{M}^{k-1}, \mathbf{z}(t_k)) \tag{5.19}$$

The expected differential entropy reduction [2] for a time interval $[t_0, t_f]$ is defined as following,

$$J = \sum_{k=1}^{f} \left\{ H(\mathbf{x}(t_k) \mid \mathbf{M}^{k-1}) - \int_{\mathcal{Z}_i} [p(\mathbf{z}(t_k) \mid \mathbf{M}^0) H(\mathbf{x}(t_k) \mid \mathbf{M}^{k-1}, \mathbf{z}(t_k))] d\mathbf{z}(t_k) \right\} \tag{5.20}$$

where $\mathbf{M}^{k-1}$ denotes the available measurement set at time t_k.

The information gain at each time step which needs to be maximized can be written as,

$$J_k = H(\mathbf{x}(t_k) \mid \mathbf{M}^{k-1}) - \int_{\mathcal{Z}_k} [p(\mathbf{z}(t_k) \mid \mathbf{M}^{k-1}) H(\mathbf{x}(t_k) \mid \mathbf{M}^{k-1}, \mathbf{z}(t_k))] d\mathbf{z}(t_k) \tag{5.21}$$

The measurement is obtained only if the target is in the FOV $\mathcal{S}(\mathbf{x}(t_k))$. Let $q(\mathbf{x}(t_k)) = P[\mathbf{x}(t_k) \in \mathcal{S}(\mathbf{x}(t_k)) \mid \mathbf{M}^{k-1}]$ indicates the probability that $\mathbf{x}(t_k)$ is in FOV $\mathcal{S}(\mathbf{x}(t_k))$. The expected differential entropy reduction at time t_k can be simplified as following,

$$J_k = q(\mathbf{x}(t_k))[H(\mathbf{x}(t_k) \mid \mathbf{M}^{k-1}) - H(\mathbf{x}(t_k) \mid \mathbf{M}^{k-1}, \mathbf{z}(t_k))] \tag{5.22}$$

For a multivariate Gaussian PDF, the differential entropy is a function of covariance matrix [22] given in the following,

$$H(\mathbf{x}) = \frac{n}{2} + \frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln[\det(\Sigma)] \tag{5.23}$$

where n is the dimension of the state vector $\mathbf{x}$.

The camera scene for the boat tracking problem is chosen to be $\mathbb{F}_p = [0, L_1] \times [0, L_2]$ where $L_1 = 960[pixel]$ and $L_2 = 540[pixel]$. The sensor FoV is modeled as a rectangle with width of $80[pixel]$ and height of $60[pixel]$. Initially, sensor locations are generated randomly within $\mathbb{F}_p$ because no information about target location is available. A optimal sensor location is selected based on prior target state PDF which maximizes the expected entropy reduction. A sensor measurement is used to update the information state of the target if the target is inside the optimal sensor FoV $\mathcal{S}$.

CHAPTER 6

SIMULATIONS AND RESULTS

In this chapter, an EER target tracking approach presented in the previous chapter is tested on a video dataset from PETS 2016 workshop. Maritime target detection results using the optical flow estimation are presented. The camera model and target model parameters estimation results are shown and discussed. Optimal sensor locations are selected to minimize the target state uncertainty. Simulation results from optimal sensor locations selection are discussed. The simulated sensor FoV is controlled such that it could track the targets in both single object tracking and multiple object tracking scenario.

6.1 Results from Optical Flow Estimation

In this section, the optical flow estimation described in the previous chapter is tested on PETS 2016 maritime video dataset. The locations for both targets are extracted at each frame based on detected bounding boxes. Bounding boxes are generated for groups of velocity vectors that reach a threshold and are close to each other. The center of the bounding box is assumed to be the location of the detected target. The detection results are demonstrated in Figure 6.1. The target trajectory is then extracted from the video dataset. The results are shown below for the leading boat target and following boat target trajectories in the region of interest in Figure 6.2 and 6.3. The noise generated by detection data compared with ground truth data is caused by dynamic background where water foams and waves have major effects.

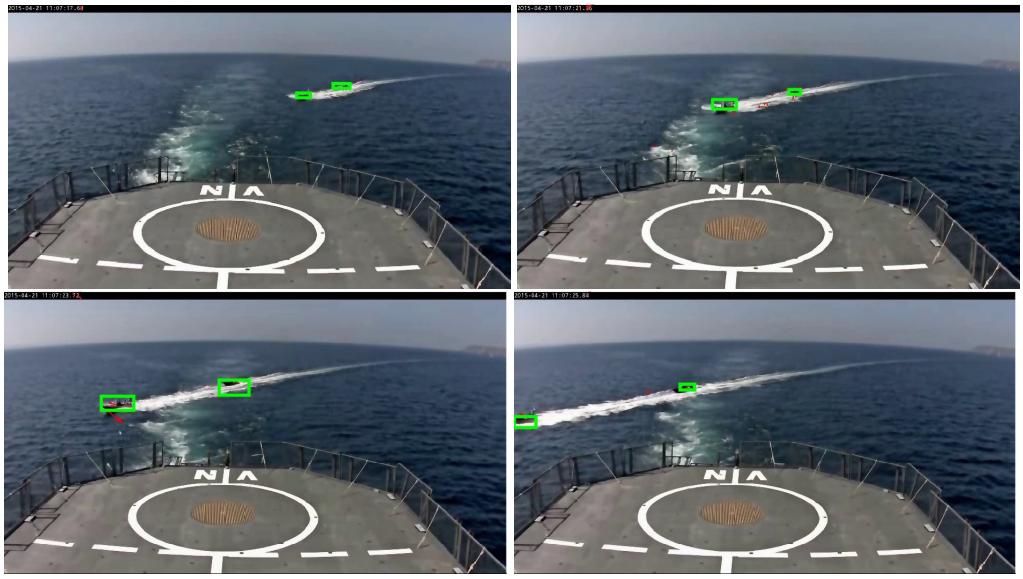


Figure 6.1: The snapshots of the video at different moments, where two speedboats were detected. Green rectangles indicate bounding boxes.

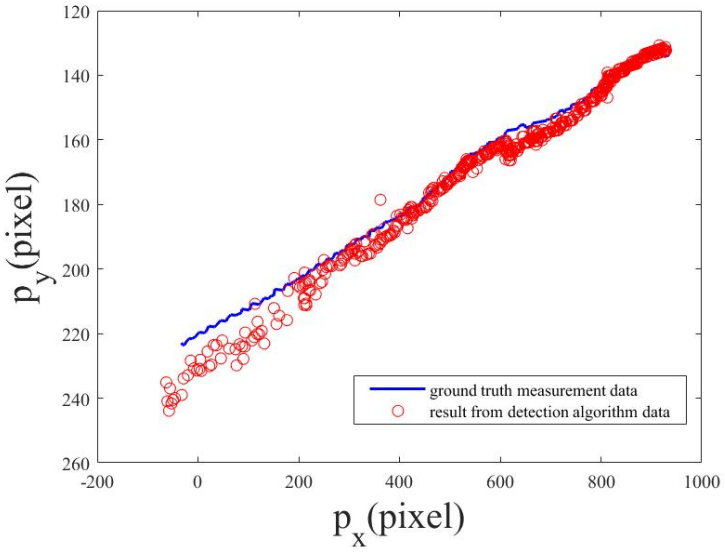


Figure 6.2: Boat trajectory from detection algorithm and ground truth data set for the leading boat.

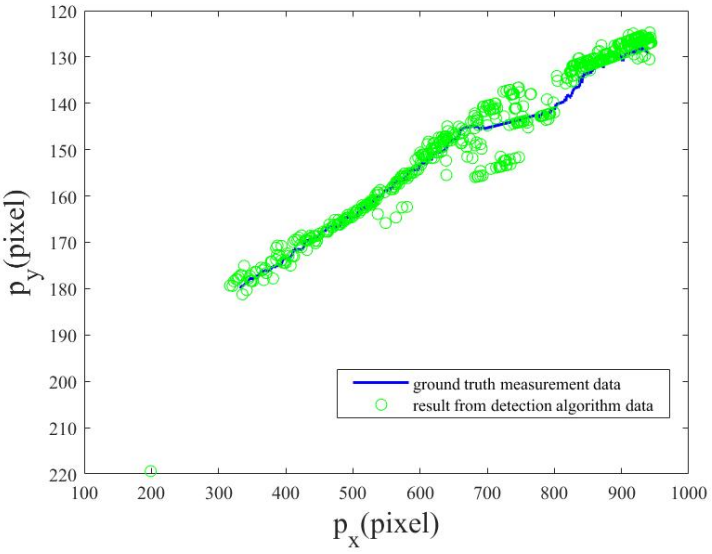
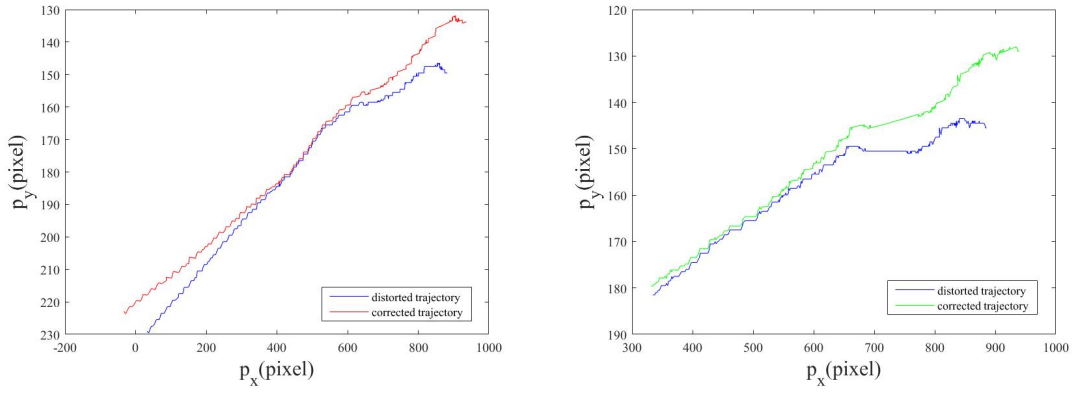


Figure 6.3: Boat trajectory from detection algorithm and ground truth data set for the following boat.

6.2 Results from Parameters Estimation

In this section, optimized camera measurement model parameters and target kinematics model parameters are presented for the leading boat target and following boat target shown in the video dataset. Since no information is available regarding the camera model and target kinematics model parameters other than the video dataset, those parameters are learned from the ground truth measurement dataset by minimizing the cost function defined in the previous chapter. Due to the image distortion effect presented in the video dataset, the first step is to correct the image distortion such that the corrected observation data can be used to estimate the parameters. Figure 6.4 shows the corrected observation dataset using image distortion parameters for both targets. As can be seen from the figure, the linearity of the corrected trajectories becomes more significant compared with original uncorrected trajectories.



(a) (b)

Figure 6.4: Results from distortion correction, (a) presents the distortion correction for the leading boat trajectory; (b) shows corrected following boat trajectory

Given the distortion corrected observation dataset, the cost function defined in the previous chapter is minimized to obtain optimized parameters for the camera measurement model and target kinematics model. Two sets of parameters are shown in Table 6.1.

Table 6.1: Optimized sets of parameters for both mobile targets

target i	$x_{init}[m]$	$y_{init}[m]$	$\theta_c[rad]$	v_c[m/s]	ψ[rad]	ϕ[rad]	α[pixel]	$z_0[m]$
1	515.54	186.85	3.56	12.85	3.63	2.97	13.5	4.5
2	591.41	200.14	3.54	12.85	3.60	2.98	13.37	4.56

Figure 6.5 and 6.6 demonstrate the predicted target trajectory against the ground truth observation data. Two sets of data are not perfectly aligned with each other because of the linear motion assumption made in the target kinematics model. However, in order to track the target using Bayesian filtering

techniques, the optimized parameters provide a reasonable initial guess for the target state such that tracking can be performed in the next step.

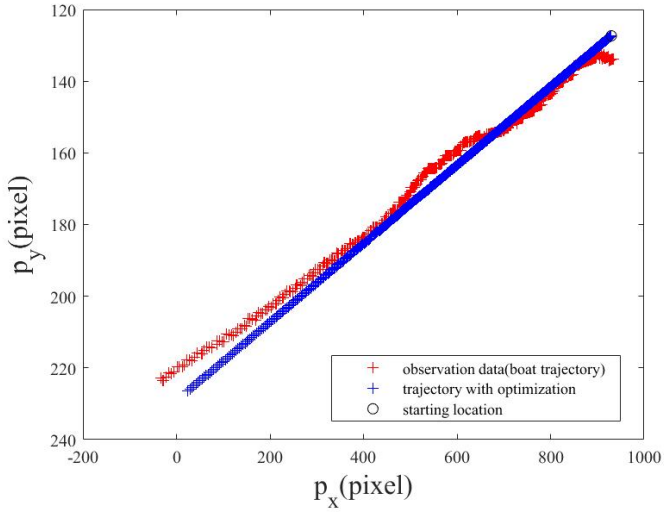


Figure 6.5: Predicted boat trajectory using corrected distortion dataset for the leading boat target.

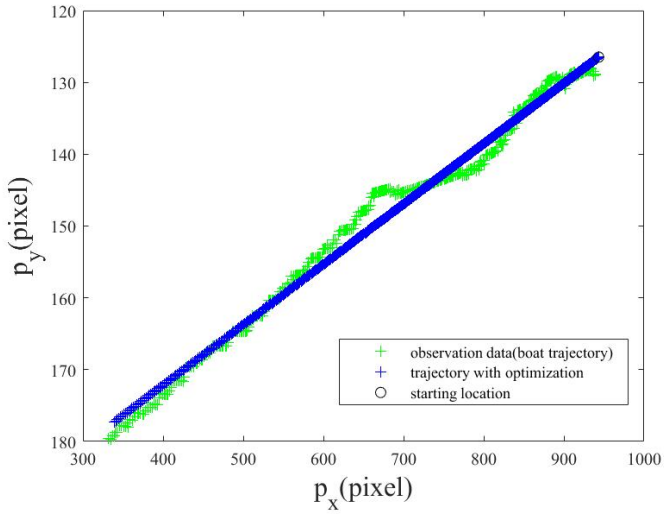


Figure 6.6: Predicted boat trajectory using corrected distortion dataset for the following boat target.

6.3 Simulations for Single Target Tracking via EER Apporach

In this section, optimal sensor location results for single target tracking are shown. A sensing control strategy based on EER is implemented such that the uncertainty presented in the target state can be minimized by placing the sensor FoV optimally. The sensor location is selected based on target prior PDF at each time step so that the location could obtain most information for planning the future measurements. The single object tracking problem is considered, and the results for the leading boat and following boat are shown separately.

The sensor trajectory and trajectory of means of target PDF are plotted in Figure 6.7. As can be seen from the plot, the sensor is able to track the target in image frame $\mathbb{F}_p$.

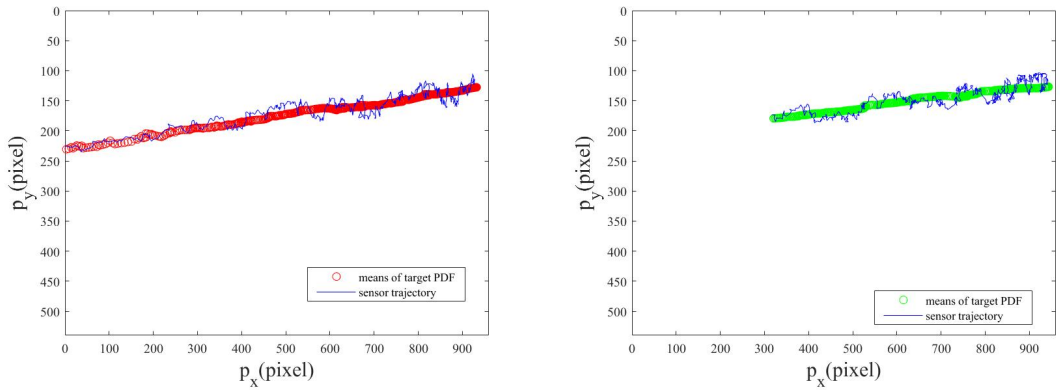
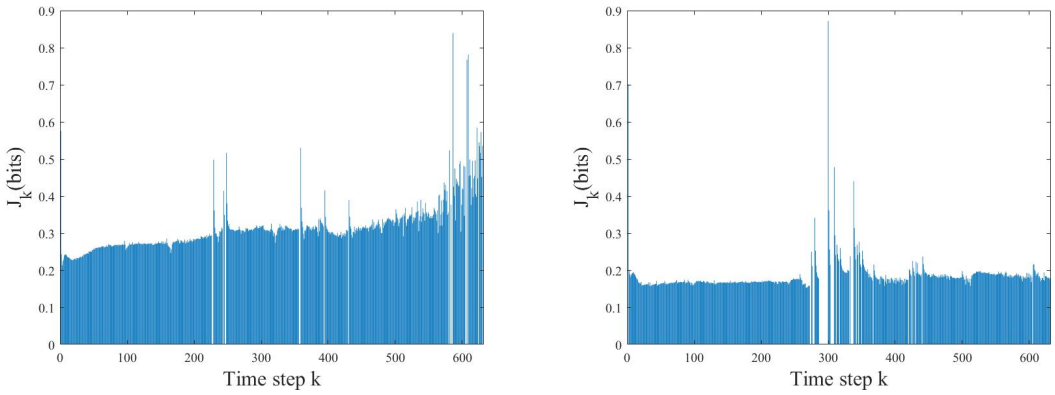


Figure 6.7: Sensor trajectory and trajectory of means of predicted target PDFs for the leading boat (a) shown in red and following boat (b) shown in green.

The information gain J_k is plotted in Figure 6.8 to show information gain at each time step. The information gain rises in the end for the leading boat, where it became less certain about the target state. This is caused by the error

presented in the measurement dataset shown in Figure 6.2 where the trajectory deviates from the ground truth dataset. However, the sensor's FoV is capable of tracking the target with small error given the dynamics of the target is linear. For the following boat, the information gain increases due to the presence of false detection and decreases to a constant level in the end.

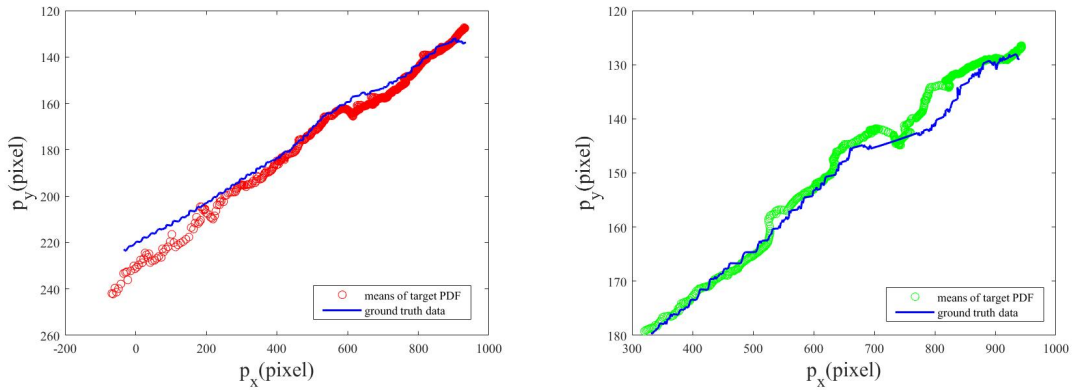


(a) (b)

Figure 6.8: Information gain at each time step for the leading target (a) and the following target (b).

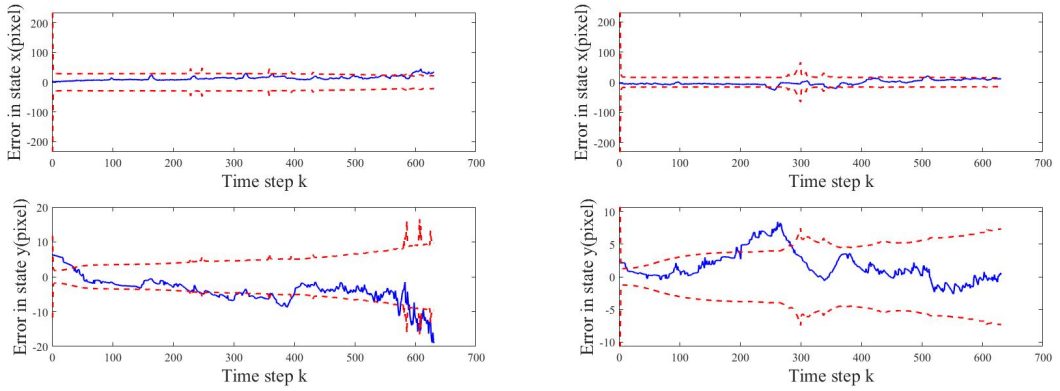
Means of posterior state estimates are plotted against the ground truth boat position in the image frame in Figure 6.9. By using the EER approach for tracking targets, the sensor can move to the optimal location at each time step where it can obtain the most information from the targets. Overall, the EER method fulfilled the requirement of tracking mobile targets separately while making the sensor location selection optimally.

State estimation error for each target is shown in Figure 6.10. The state estimation error is calculated within the image plane where means of posterior distribution of state x and y are projected onto image plane $\mathbb{F}_p$. The state estimation error in state y of the leading target increases because of the detection errors presented in the measurement dataset.



(a) (b)

Figure 6.9: Means of posterior state estimation in image frame for leading boat (a) and following boat (b).



(a) (b)

Figure 6.10: State estimation error compared with ground truth for leading boat (a) and following boat (b) where red line represented 3 sigma bound.

The posterior target PDF at different time step and corresponding selected FOV location are showed in Figure 6.11 and 6.12. The sensor's FoV is controlled to obtain information from prior distribution for each target, and it can track the mobile object sufficiently.

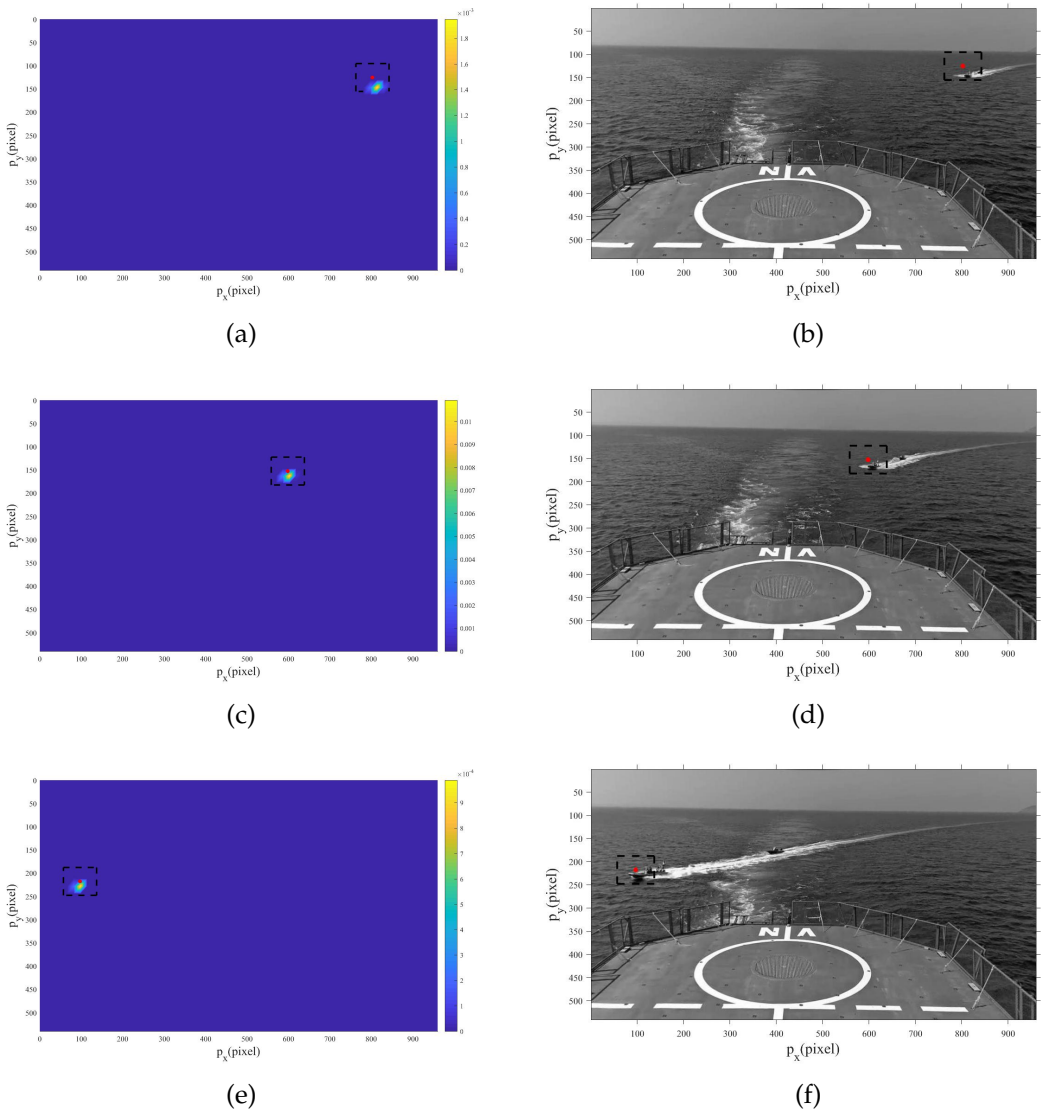


Figure 6.11: The posterior state PDFs and selected sensor FOV at different time steps for the leading boat are shown in (a), (c), (e). Corresponding original image frames with selected FoV are shown in (b), (d), (f).

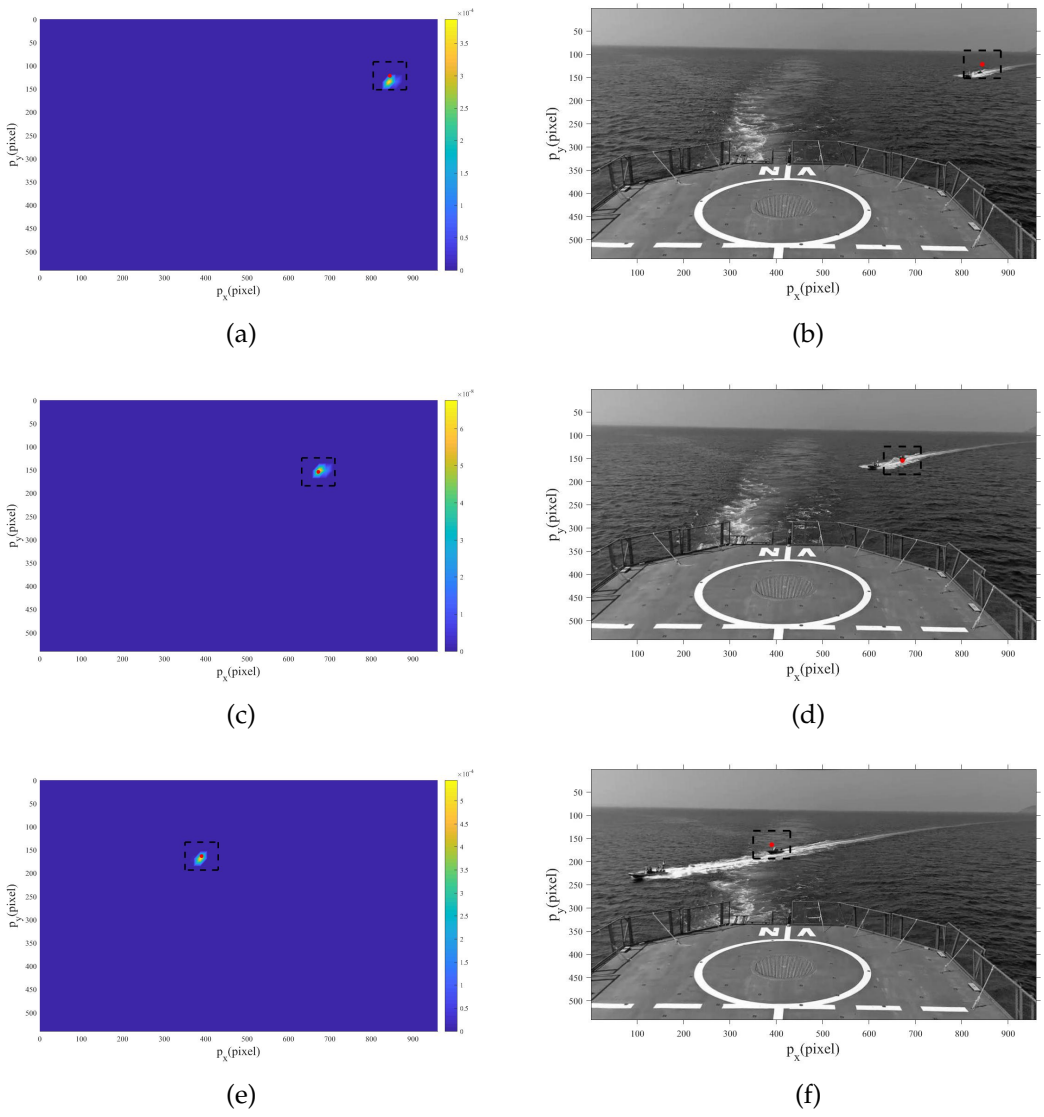


Figure 6.12: The posterior state PDFs and selected sensor FOV at different time steps for the following boat are shown in (a), (c), (e). Corresponding original image frames with selected FoV are shown in (b), (d), (f).

6.4 Simulations for Multiple Target Tracking via EER Approach

Simulation results for tracking both targets are presented in this section. The simulated FoV can choose one target to observe at each time step k. The EER approach allows the sensor's FoV to be selected for the target that obtains the most information.

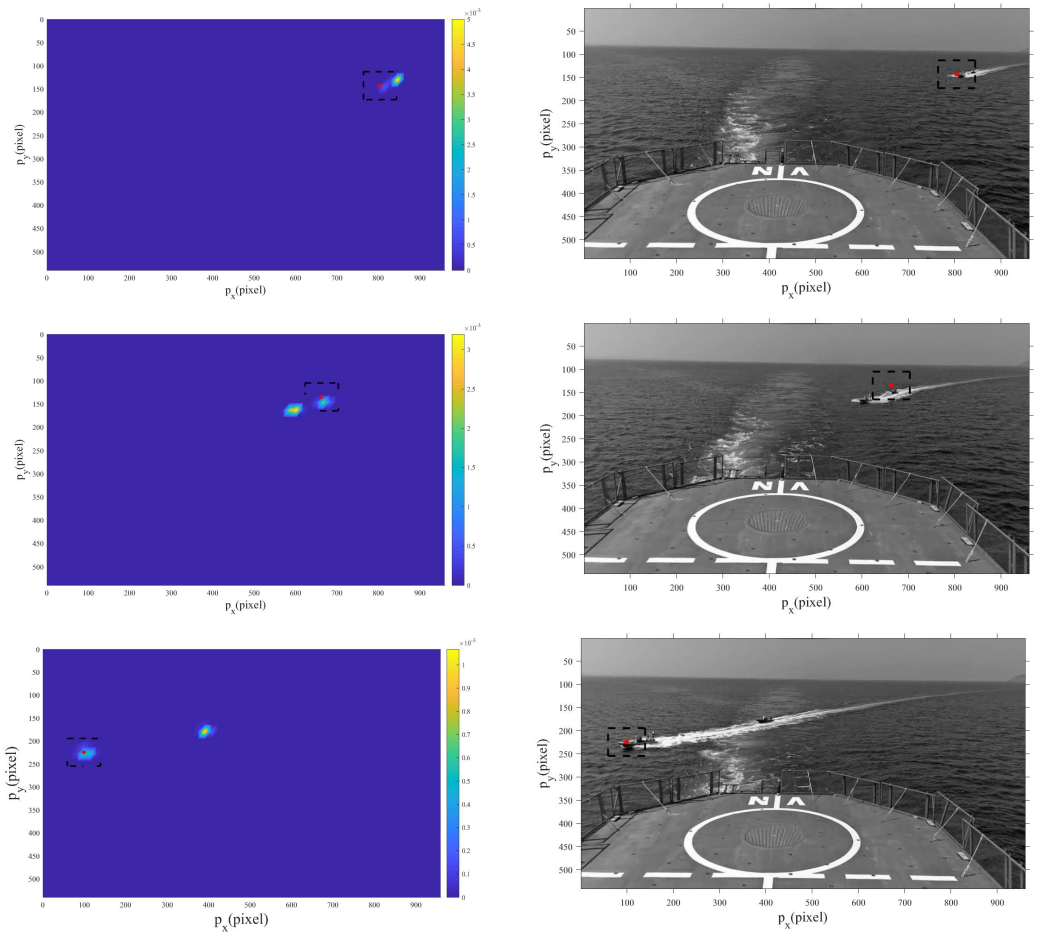
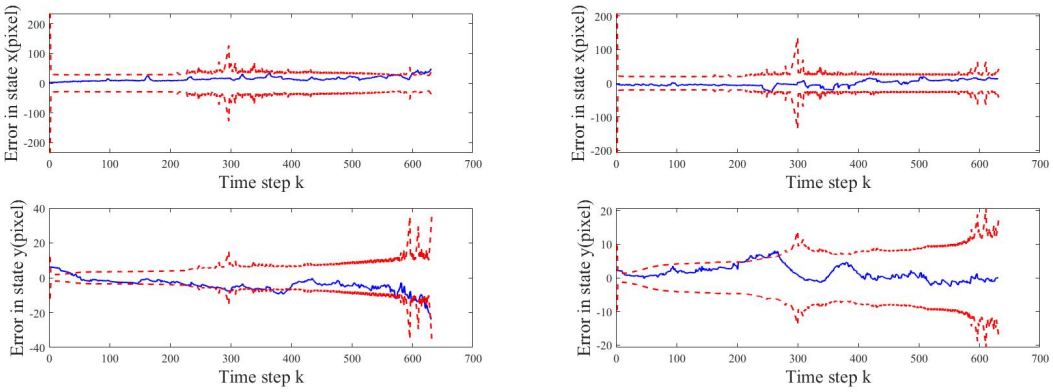


Figure 6.13: The posterior state PDFs and selected sensor FOV at different time steps for the both targets are shown in the first column. Corresponding original image frame with selected FoV are shown in the second column.

As shown from Figure 6.13, the FoV location is controlled at each time step to select the target that is less certain in terms of its PDF compared with the other target and thus obtain more information. The simulated FoV would switch between two targets over time such that both targets can be tracked.

The state estimation error for both targets is shown in Figure 6.14. For the leading target, the estimation error presented in state y increased due to the detection accuracy. The spikes presented in 3σ bound for one target indicate that the simulated FoV is observing the other target. Overall, the purposed method has demonstrated the effectiveness in tracking both targets while maintaining low estimation errors.



(a) (b)

Figure 6.14: State estimation error compared with ground truth for leading boat (a) and following boat (b) where red line represented 3 sigma bound.

CHAPTER 7

CONCLUSION AND FUTURE WORK

This thesis introduces a mobile target detection and tracking algorithm with a sensor planning strategy via expected entropy reduction (EER) to minimize the target state's uncertainty in the maritime environment. First, an optical flow estimation is used for detecting mobile targets under the maritime environment. Despite the dynamic background of the maritime environment and the noise presented in the video sequence, the proposed method can detect moving targets with a small number of false detection. Secondly, a cost function is designed so that the proposed kinematics parameters can be estimated and optimized from the ground truth measurement dataset. Thirdly, a Bayesian approach based on extended Kalman filter is employed for tracking detected targets. A sensor planning strategy based on expected entropy reduction (EER) is developed so that the optimal sensor location is selected to minimize the uncertainty of the target states.

Future work should focus on improving detection accuracy given a dynamic maritime environment. Currently, false detection would occur when wakes generated by the mobile target are highly dynamic. Dynamic background models that can precisely separate the wakes or waves in the background from the targets should be studied. In this thesis, a known number of mobile targets is assumed for tracking purposes while in future work an unknown number of targets should be considered. The proposed method should be compared with other available methods on a broader range of maritime datasets.

BIBLIOGRAPHY

[1] Tom Cane and James Ferryman. Saliency-based detection for maritime object tracking. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 18–25, 2016.

[2] Silvia Ferrari and Thomas A. Wettergren. *Information-Driven Planning and Control*. MIT Press, 2021. Available at `https://mitpress.mit.edu/contributors/silvia-ferrari`.

[3] Hongchuan Wei and Silvia Ferrari. A geometric transversals approach to sensor motion planning for tracking maneuvering targets. *IEEE Transactions on Automatic Control*, 60(10):2773–2778, 2015.

[4] Pingping Zhu, Silvia Ferrari, Julian Morelli, Richard Linares, and Bryce Doerr. Scalable gas sensing, mapping, and path planning via decentralized hilbert maps. *Sensors*, 19(7):1524, 2019.

[5] Jake R Gemerek, Silvia Ferrari, and John D Albertson. Fugitive gas emission rate estimation using multiple heterogeneous mobile sensors. In *2017 ISOCS/IEEE International Symposium on Olfaction and Electronic Nose (ISOEN)*, pages 1–3. IEEE, 2017.

[6] Guoxian Zhang, Silvia Ferrari, and M Qian. An information roadmap method for robotic sensor path planning. *Journal of Intelligent and Robotic Systems*, 56(1-2):69–98, 2009.

[7] 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(8):1993–2016, 2017.

[8] Jing Zhong et al. Segmenting foreground objects from a dynamic textured background via a robust kalman filter. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 44–50. IEEE, 2003.

[9] Dilip K Prasad, C Krishna Prasath, Deepu Rajan, Lily Rachmawati, Eshan Rajabaly, and Chai Quek. Challenges in video based object detection in maritime scenario using computer vision. *arXiv preprint arXiv:1608.01079*, 2016.

[10] Vitaly Ablavsky. Background models for tracking objects in water. In *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, volume 3, pages III–125. IEEE, 2003.

[11] Sriram Varadarajan, Paul Miller, and Huiyu Zhou. Region-based mixture of gaussians modelling for foreground detection in dynamic scenes. *Pattern Recognition*, 48(11):3488–3503, 2015.

[12] Michael Gregory Ross. *Exploiting texture-motion duality in optical flow and image segmentation*. PhD thesis, Massachusetts Institute of Technology, 2000.

[13] John L Barron, David J Fleet, and Steven S Beauchemin. Performance of optical flow techniques. *International journal of computer vision*, 12(1):43–77, 1994.

[14] Yaakov Bar-Shalom, X Rong Li, and Thiagalingam Kirubarajan. *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

[15] Fredrik Gustafsson, Fredrik Gunnarsson, Niclas Bergman, Urban Forssell, Jonas Jansson, Rickard Karlsson, and P-J Nordlund. Particle filters for positioning, navigation, and tracking. *IEEE Transactions on signal processing*, 50(2):425–437, 2002.

[16] Nathan Funk. A study of the kalman filter applied to visual tracking. *University of Alberta, Project for CMPUT*, 652(6), 2003.

[17] Domenico Bloisi and Luca Iocchi. Argos—a video surveillance system for boat traffic monitoring in venice. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(07):1477–1502, 2009.

[18] Hongshan Yu, Yaonan Wang, Fei Kuang, and Qin Wan. Multi-moving targets detecting and tracking in a surveillance system. In *Fifth World Congress on Intelligent Control and Automation (IEEE Cat. No. 04EX788)*, volume 6, pages 5253–5257. IEEE, 2004.

[19] Keith LeGrand and Silvia Ferrari. Exploiting bounded sensor field-of-view geometry in tracking and sensor planning problems. *arXiv preprint arXiv:2004.00795*, 2020.

[20] W Lu, G Zhang, and Silvia Ferrari. A comparison of information theoretic

functions for tracking maneuvering targets. In *2012 IEEE Statistical Signal Processing Workshop (SSP)*, pages 149–152. IEEE, 2012.

[21] Feng Zhao, Jaewon Shin, and James Reich. Information-driven dynamic sensor collaboration. *IEEE Signal processing magazine*, 19(2):61–72, 2002.

[22] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[23] Roy De Maesschalck, Delphine Jouan-Rimbaud, and Désiré L Massart. The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, 50(1):1–18, 2000.

[24] James Manyika and Hugh Durrant-Whyte. *Data fusion and sensor management: a decentralized information-theoretic approach*. Prentice Hall PTR, 1995.

[25] Hongchuan Wei, Wenjie Lu, and Silvia Ferrari. An information value function for nonparametric gaussian processes. *arXiv preprint arXiv:1406.3296*, 2014.

[26] Guoxian Zhang, Silvia Ferrari, and Chenghui Cai. A comparison of information functions and search strategies for sensor planning in target classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(1):2–16, 2011.

[27] Hongchuan Wei, Wenjie Lu, Pingping Zhu, Silvia Ferrari, Robert H Klein, Shayegan Omidshafiei, and Jonathan P How. Camera control for learning nonlinear target dynamics via bayesian nonparametric dirichlet-process gaussian-process (dp-gp) models. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 95–102. IEEE, 2014.

[28] Nicholas R Gans, Guoqiang Hu, and Warren E Dixon. Keeping multiple objects in the field of view of a single ptz camera. In *2009 American Control Conference*, pages 5259–5264. IEEE, 2009.

[29] Luis Patino, Tom Cane, Alain Vallee, and James Ferryman. Pets 2016: Dataset and challenge. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–8, 2016.

[30] Dhara Patel and Saurabh Upadhyay. Optical flow measurement using lucas kanade method. *International Journal of Computer Applications*, 61(10):6–10, 2013.

[31] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*, pages 839–846. IEEE, 1998.

[32] Daniel Santana-Cedrés, Luis Gomez, Miguel Alemán-Flores, Agustín Salgado, Julio Esclarín, Luis Mazorra, and Luis Alvarez. An iterative optimization algorithm for lens distortion correction using two-parameter models. *Image Processing On Line*, 6:326–364, 2016.

[33] Luis Alvarez, Luis Gómez, and J Rafael Sendra. An algebraic approach to lens distortion by line rectification. *Journal of Mathematical Imaging and Vision*, 35(1):36–50, 2009.

[34] Hongchuan Wei, Pingping Zhu, Miao Liu, Jonathan P How, and Silvia Ferrari. Automatic pan–tilt camera control for learning dirichlet process gaussian process (dpgp) mixture models of multiple moving targets. *IEEE Transactions on Automatic Control*, 64(1):159–173, 2018.

[35] Kenji Hata and Silvio Savarese. Cs231a course notes 1: Camera models. Available at `https://web.stanford.edu/class/cs231a/course_notes/01-camera-models.pdf`.

[36] Simo Särkkä. *Bayesian filtering and smoothing*, volume 3. Cambridge University Press, 2013.

[37] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic robotics*, volume 1. MIT press Cambridge, 2000.