Ocean Engineering 252 (2022) 111168

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Heterogeneous multi-sensor tracking for an autonomous surface vehicle in a littoral environment☆

Øystein Kaarstad Helgesen *, Kjetil Vasstein, Edmund Førland Brekke, Annette Stahl

Norwegian University of Science and Technology, Department of Engineering Cybernetics, O. S. Bragstads plass 2D, Trondheim, 7034, Trøndelag, Norway

ARTICLE INFO

Keywords:
Sensor fusion
Autonomy
Target tracking
Maritime
Situational awareness
Open dataset

ABSTRACT

Sensor fusion plays a key part in autonomous surface vehicles, however, the high cost of sensors makes the barrier of entry in this research field quite high. In this work, we present a complete system for sensor fusion on the milliAmpere autonomous ferry research platform as well as an open sensor fusion dataset for maritime tracking across two environments. Individual sensors and their detection pipelines are evaluated across various detection metrics. We also evaluate the tracking performance of the sensors both individually and in fusion using a multi-sensor extension of the JIPDA multi-target tracker. We find that the different environments have distinct challenges precluding the use of only a single sensor. Utilizing multiple sensors, either individually or in fusion, can mitigate these issues increasing the safety margins of the situational awareness system.

1. Introduction

Interest in autonomous surface vehicles (ASV) such as ferries has increased rapidly in the last few years. The potential for safer operations, reduced cost, and increased availability compared to human-operated vessels have been key factors behind this interest. Compared to land-based autonomous vehicles such as cars or trucks the environment an ASV operates in is comparatively simpler with more open areas and fewer obstacles and targets for the vehicle to navigate around. Maritime autonomous operations also have the option of onshore monitoring and control centers that could take control of an ASV remotely when needed which should enable more rapid deployment compared to land-based vehicles.

A key part of autonomous operations is situational awareness. Without an accurate understanding of the position and path of other vessels, both path planning and collision avoidance would be impacted, reducing the safety margins of the vessel. This situational awareness system processes raw sensor data from one or more sensors, extracts useful information in the form of detections, and estimates the current and future states of the vessels using target tracking.

Target tracking is a form of state estimation used when multiple measurements can be received, requiring the estimator to evaluate if a measurement originates from a target or from false alarms. These tracking methods exist in multiple forms. Single target trackers such as the Probabilistic Data Association Filter (PDAF) and derivatives such as the Integrated PDAF (IPDA) (Bar-Shalom and Tse, 1975; Musicki et al., 1992) assume only a single target exists in the surveillance region. Each individual measurement is used to update the state estimate based on the likelihood of it originating from the target or from false alarms. If multiple targets are presented one option is to instance individual single-target trackers for each target. For well-separated targets this is a valid solution, however, when the targets operate closer together such as in urban environments this will yield sub-optimal performance, requiring multi-target tracking.

The key difference in multi-target tracking is the possibility of a measurement originating from other targets as well as from false alarms. Multiple methods have been developed to deal with the multi-tracking problem. Joint PDAF (Fortmann et al., 1980) and Joint IPDA (Musicki and Evans, 2004) are multi-target extensions of the PDAF based tracking methods allowing the data association process to consider multiple targets for a measurement's origin. Other multi-target tracking methods based on multiple hypotheses such as the track-oriented multiple hypothesis tracker (TOMHT) also exist (Bar-Shalom et al., 2002) as an option to the PDAF based trackers.

In the automotive field, several studies have been performed on detection and tracking with sensor fusion (Cho et al., 2014; Chavez-Garcia and Aycard, 2016), however, the maritime domain has received less attention. Several works examine tracking with single sensors such as radars (Schuster et al., 2014; Wilthil et al., 2017; Fowdur et al., 2021) and cameras (Schöller et al., 2020; Wolf et al., 2010), however, multi-sensor tracking is not addressed. Haghbayan et al. (2018) presented

☆ The datasets used in this work are available at https://autoferry.github.io/sensor_fusion_dataset/.

* Corresponding author.
E-mail addresses: oystein.k.helgesen@ntnu.no (Ø.K. Helgesen), kjetil.vasstein@ntnu.no (K. Vasstein), edmund.brekke@ntnu.no (E.F. Brekke), annette.stahl@ntnu.no (A. Stahl).

https://doi.org/10.1016/j.oceaneng.2022.111168
Received 19 October 2021; Received in revised form 7 February 2022; Accepted 22 March 2022
Available online 6 April 2022

a detection framework using sensor fusion in the maritime domain, however, no tracking was performed. Cormack et al. (2020) fused radar with infrared cameras to perform multi-target tracking. Radar was also fused with AIS in Gaglione et al. (2018) using belief propagation based sensor fusion. A preliminary exploration of multi-sensor, multi-target tracking was performed by Helgesen et al. (2019) using a small dataset from a land-based sensor rig.

While several studies on multi-sensor fusion have been performed, there is in general a lack of open datasets for maritime situational awareness and multi-target tracking. In the automotive world, several datasets exist such as KITTI (Geiger et al., 2012). The maritime domain is currently lacking this range, however recent contributions do seem to indicate this is changing. Cheng et al. (2021) published a multi-sensor dataset for stereo vision, simultaneous localization and mapping, and water segmentation for ASVs while Brekke et al. (2020) included an open dataset for multi-target tracking with radar measurements. The data used in Fowdur et al. (2021) are also available for NATO or EU member states but only upon request. There is however to the authors' knowledge no publicly available multi-sensor tracking datasets available, increasing the barrier of entry in this field.

Several contributions are presented in this work. A multi-sensor extension of the JIPDA is detailed and validated experimentally on a large benchmark dataset with four heterogeneous sensors and multiple targets with ground truth. This dataset includes radar, lidar, infrared and electro-optical detections from a vessel-mounted sensor rig in two distinct environments. The dataset is made publicly available (see link in the footnote), we also aim to publish the raw sensor data subject to privacy concerns. The autonomous platform and sensor pipelines are also explored in detail and evaluated using a range of performance metrics.

1.1. Notation

The following notation standards are used throughout this work:

- **t**: lower case bold letters represent vectors.
- **R**: upper case bold letters represent matrices.
- T/T/t/t: non-bold letters in both cases, both italic and non-italic, represent single values.

2. Sensor platform

milliAmpere, Fig. 1, is an urban autonomous passenger ferry prototype/research platform developed at the Norwegian University of Science and Technology (NTNU). As a research platform, milliAmpere is equipped with a complete navigation and sensing system designed to make future all-autonomous operations possible. Two battery-powered azimuth thrusters allow for rapid course changes due to the small size of the craft (cf. Fig. 2). Top speed is limited at around 5 knots due to regulatory reasons which is slower than most similarly sized vessels.

2.1. Navigation system

Accurate positioning is a key component in autonomous operations. Many of the ownship modules utilize position information, including both docking and situational awareness systems, where increased accuracy can yield higher safety margins. A dual antenna Hemisphere Vector VS330 GNSS compass provides both position and heading information. Real-Time Kinematic (RTK) corrections are transmitted to the ferry from an antenna situated at the NTNU Gløshaugen campus yielding a manufacturer specified position accuracy of 10 mm and heading accuracy of 0.05°.

Fig. 1. *milliAmpere*, an autonomous ferry research platform.

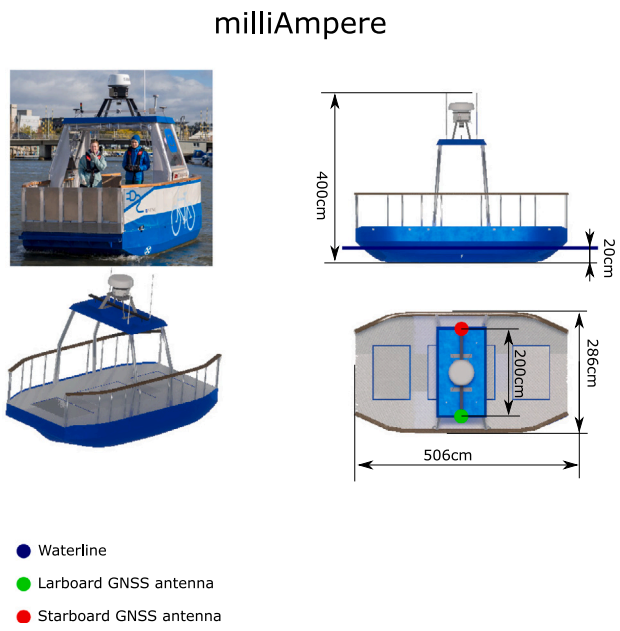


Fig. 2. Dimensions of *milliAmpere* as depicted in Vasstein (2021).

2.2. Sensing system

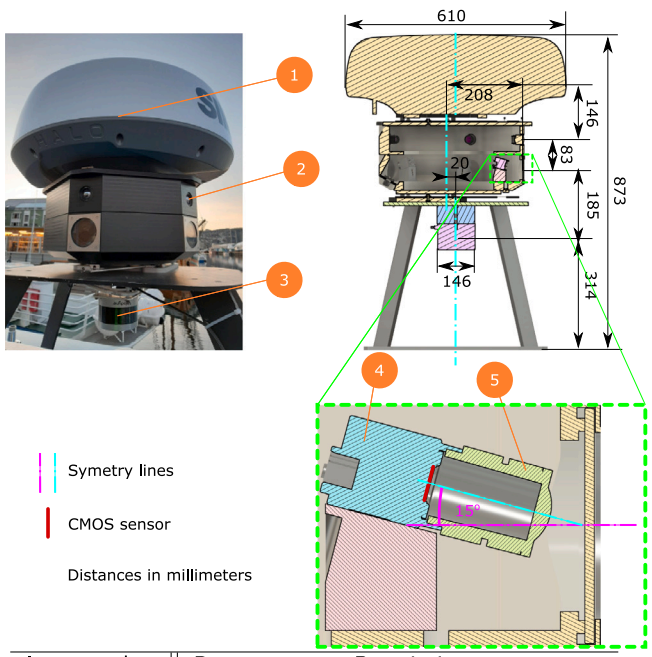
With 10 cameras, a radar, and a lidar, *milliAmpere* is highly suited for the collection of datasets for heterogeneous sensor fusion. Minimal sensor obstruction is achieved by mounting the sensor on top of the ferry as shown in Fig. 3.

2.2.1. Lidar

Lidar data comes from a Velodyne VLP-16 with a maximum range of 100 m. 16 rotating laser beams provide a 360x30° field of view 3D point cloud at a rate of about 10 Hz. This sensor provides highly accurate range estimates at ±3 cm. However, the low vertical angular resolution at 2.0° does pose challenges at the far end of its operational interval where the gap between beams can be as large as 2.8 m (80 m range). Each point cloud contains a series of points with the following data format: $\mathbf{p}_i = [p_x, p_y, p_z, i]$ where $p_{x/y/z}$ is the Cartesian coordinates of the return and i the intensity of the return signal.

2.2.2. Radar

Radars have traditionally been the most common sensor system in maritime applications due to their reliability and range. *milliAmpere*



Item number	Part	Description
1	Radar	Simrad 4G
2	IR camera	FLIR Boson 640
3	Lidar	Velodyne Puck | VLP-16
4	Image sensor	FLIR Blackfly GigE
5	Camera lens	Kowa LM6JC

Fig. 3. Sensor system dimensions as depicted in Vasstein (2021).

is equipped with a Simrad Broadband 4G radar which has a built-in detection system yielding sensor data in the form of radar spokes. Each spoke is composed of an array of range cells with binary detection values as well as the associated azimuth angle, θ, giving a series of polar detections of the form $\mathbf{z}_i = \begin{bmatrix} r_i \\ \theta_i \end{bmatrix}$. Rotating at 48RPM/0.8 Hz the radar is far slower than the lidar and has a much lower angular accuracy due to its 5.2° beamwidth. The radar was manually tuned using a radar plotter.

2.2.3. Cameras

The advent of cheap, performant computing hardware and deep learning based detection methods has made cameras a highly relevant sensor in autonomous operations. *milliAmpere* is equipped with a 360° camera rig containing 5 electro-optical and 5 infrared cameras. Electro-optical imaging data comes from FLIR Blackfly S 50-S5C cameras equipped with a 6 mm lens yielding a horizontal field of view of 77.8°. With 5 cameras this yields some overlap between the cameras ensuring complete 360° coverage. Each camera is connected to a single Nvidia Jetson TX2 board embedded in the sensor rig using Power-over-Ethernet. Due to the bandwidth constraints of the single gigabit link out of the sensor rig each camera operates at half resolution, 1224x1020, and at a rate of 5 Hz. Sensor data are transmitted over ethernet in raw Bayer format. Once received, images are debayered to generate color images and then undistorted to correct for lens distortions.

Infrared imaging data are supplied by 5 FLIR Boson 640 cameras with a 4.9 mm lens and a horizontal field of view of 95 degrees. Using an uncooled vanadium-oxide microbolometer sensor the cameras provide infrared images at a resolution of 640x512 over USB. Operating at 9 Hz the cameras are sensitive to long-wave infrared radiation with a spectral band of 7.5 μm-13.5 μm. Two USB hubs connect all 5 cameras to the TX2 board which transmits the images over ethernet. Images are then undistorted before being passed on to the detection pipeline. Both cameras are calibrated according to Zhang (2000) to find their intrinsic parameters.

3. Detection pipeline

Sensor data usually requires one or more processing steps for it to be interpretable by an ASVs situational awareness system. This section presents the detection pipelines used to extract useful information from raw sensor data provided by the ASVs sensor systems.

3.1. Radar/Lidar pipeline

Both active sensors, the radar and lidar, will yield positive returns for land and buildings if within range. Further processing of the sensor data is therefore required to avoid issues in the tracking system. Once converted to a point cloud, both the radar and lidar share the same sensor data format allowing the remaining detection pipeline to be shared. This pipeline was first introduced in Wilthil et al. (2017) and has later been extended to support lidar point clouds as well.

3.1.1. Radar processing

While the radar already supplies us with detections in the form of spokes most methods for target tracking assume that a target can generate at most one measurement per time step. Raw detection data from the radar does not meet this assumption, any target with radar returns will almost always generate detections in multiple range cells and radar spokes.

The first step in the detection pipeline is to convert positive detections from a single spoke from polar to Cartesian coordinates in the radar frame (r) according to

$$\mathbf{z}_i^r = \begin{bmatrix} r_i \cos\theta_i \\ r_i \sin\theta_i \\ 0 \end{bmatrix} \tag{1}$$

where r_i is the range of the detection given by the range cell and θ_i the azimuth angle of the spoke i. Using data from the navigation system each detection is then transformed into a world-fixed local North-East-Down (NED) frame (w) by a rotation $\mathbf{R}_r^w$ and translation $\mathbf{t}_r^w$ from the radar frame according to data supplied by the navigation system.

$$\mathbf{z}_i^w = \mathbf{R}_r^w \mathbf{z}_i^r + \mathbf{t}_r^w \tag{2}$$

Converted detections from a single radar rotation are then aggregated into a 2D point cloud of positive detections fixed on the ocean surface for further processing.

3.1.2. Lidar processing

Sensor data from the lidar is natively supplied as point clouds, however, a conversion from the sensor frame to a world-fixed NED frame is still required for further processing. For any point i with coordinates $\mathbf{p}_i^l$ in the lidar frame l the conversion is

$$\mathbf{p}_i^w = \mathbf{R}_l^w \mathbf{p}_i^l + \mathbf{t}_l^w. \tag{3}$$

3.1.3. Land returns

To avoid land returns from inducing false tracks in the tracking system the point clouds are processed according to a land map supplied by the Norwegian Mapping Authority. A binary occupancy grid centered on the local NED origin is generated and projected onto the land map. Each cell overlapping with land is given a positive occupancy value while cells covering sea will receive a negative value. To compensate for potential uncertainties in the sensor data transforms and mapping data the land area is also dilated by a fixed amount.

Additional masking is often required due to a lack of certain features in the map data such as docks. This is illustrated in Fig. 4 where the dilated land map does not cover all land returns from the radar. Using freely available online tools supplementary GeoJSON sensor masks were manually drawn to cover these features. Combining the land map with sensor masks ideally yields a point cloud with zero land returns when any points landing in occupied cells are removed.

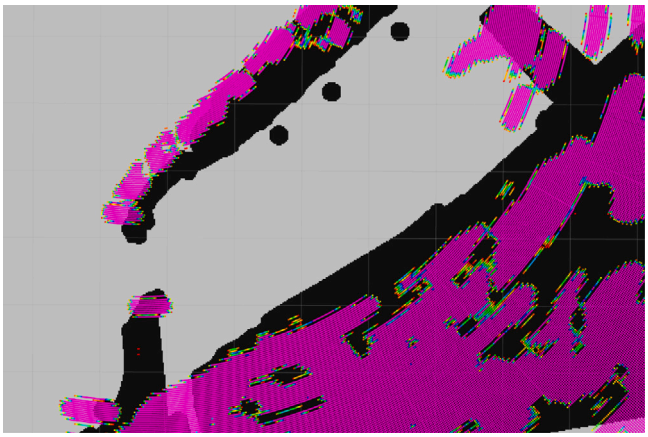


Fig. 4. Radar point cloud overlaid on dilated land map. Black fields are land areas, gray areas ocean. Radar point cloud is shown as colored squares where color is given by return intensity.

3.1.4. Clustering

While map-based processing does remove many potential false tracks a single target can still yield multiple returns due to the resolution of the sensor. For the assumptions of the tracking system to hold these need to be merged. This is done by clustering the land filtered point cloud. By assuming that closely spaced detections originate from the same target the clustering algorithm groups these together into a cluster. The centroid of this cluster is then assumed to be the true detection. Points are added to the cluster if the Euclidean distance from point i to any point j in the cluster is less than a specified threshold T:

$$\|\mathbf{p}_i - \mathbf{p}_j\| \leq T. \tag{4}$$

The clustering algorithm is an evolution of the method described in Wilthil et al. (2017) and is based on single-link hierarchical clustering using k-d trees for optimization purposes. A minimum cluster size of 5 points is also enforced to minimize potential false alarms from wakes.

3.2. Camera pipeline

As passive sensors, both the infrared and electro-optical cameras, require different detection pipelines than the active sensors. The lack of native range information in the sensor data requires more complex detection methods than simple clustering to extract useful information for the sensor fusion system to utilize.

3.2.1. Image detection

Yolo v4 (Bochkovskiy et al., 2020), a state-of-the-art object detection method based on the Yolo family of detectors, is utilized in this work to process image data. While EfficientDet (Tan et al., 2020) can yield greater accuracy for its largest models, Yolo v4 achieves similar results to the smaller models while running at roughly twice the speed. Yolo v4 is a single-stage detector as opposed to a two-stage detector. Two-stage detectors such as Faster R-CNN (Ren et al., 2015) use a region proposal method to find objects proposals which are then processed. Single-stage detectors on the other hand skip this part utilizing a sampling of pre-specified locations resulting in reduced computational complexity and greater speeds. The output of this detection process is a prediction containing one or more bounding boxes (pixel coordinates) and their associated classification and confidence score. The detector was trained on the COCO dataset (Lin et al., 2014) without target-domain specific data. Fine tuning pre-trained models on domain specific data has been shown to yield minor improvements in detection performance and robustness when evaluated on a benchmark dataset from a similar domain (Landsnes, 2021). The lack of infrared specific training data is a greater issue impacting the detection performance of the infrared sensors to a larger degree. The use of deep learning based transformers (Kniaz et al., 2018; Özkanoğlu and Ozer, 2022) on existing color image datasets to approximate their infrared equivalents is a possible solution to this issue that does not require the collection and labeling of large infrared datasets.

3.2.2. Bounding box processing

Images from each type of camera are processed individually by the detector. For each bounding box corresponding to an object of interest (boat, kayak, etc.) a measurement is generated from the object's pixel coordinates and the camera's intrinsic and extrinsic calibration parameters. Due to the passive nature of the cameras, no explicit range information is given by the sensor data, the bearing of the target is, therefore, the most natural measurement to extract. Given a bounding box a for camera b we assume the true center of the target is given by the bottom center of the predicted bounding box, $\mathbf{p}_b^a = [p_x, p_y]$, which should in the ideal case correspond with the waterline of the target. In the camera coordinate system, c, the bearing measurement can be estimated based on the cameras focal length, $\mathbf{f} = [f_x, f_y]$ and principal point, $\mathbf{p}_p = [c_x, c_y]$, in the camera coordinate system according to

$$\theta_a^b = \frac{p_x - c_x}{w^b} FOV_x^b \tag{5}$$

where w^b is the image width in pixels. FOV_x^b is the horizontal field of view of the camera b and is given by

$$\mathbf{FOV}^b = \begin{bmatrix} FOV_x^b \\ FOV_y^b \end{bmatrix} = \begin{bmatrix} 2\arctan\frac{w^b}{2f_x^b} \\ 2\arctan\frac{h^b}{2f_y^b} \end{bmatrix} \tag{6}$$

where h^b is the image height in pixels. Once this process is completed for all bounding boxes in the image the measurements are then converted to a vessel fixed local NED frame. Due to the cameras being mounted approximately in the center of the vessel we only need to compensate for a rotation ϕ_k, given by

$$\phi_k = \psi_k + \varphi_{vc}^b \tag{7}$$

where ψ_k is the rotation of milliAmpere relative to a world-fixed NED frame at time k given by the navigation system and φ_{vc}^b the fixed rotation between vessel center (vc) and camera b. This results in the vessel-fixed (v) NED measurement

$$\theta_a^v = \theta_a^b + \phi_k. \tag{8}$$

Once images from all cameras of a single type have been processed the bearing measurements are then aggregated to form a single measurement vector in the vessel NED frame.

4. Multi-sensor JIPDA

The Visibility Interactive Multiple Model JIPDA (VIMMJIPDA) of Brekke et al. (2021) is a modern formulation of the Markov-Chain Two JIPDA. Target motion is described using multiple models such as the constant velocity (CV), constant turn rate (CT) and constant acceleration (CA) model to improve estimation accuracy. A special case of this formulation utilizing a single CV motion model is used in this work which has been extended to support multi-sensor fusion. In this section we introduce the CV model as well as the key changes made to support multiple sensors, leaving the complete derivation to Brekke et al. (2021).

4.1. Motion model

In the CV motion model, target states are described by their position and velocity, $\mathbf{x} = [p_x, p_y, v_x, v_y]$, in a north-east-down (NED) reference frame. In continuous-time the model is described by

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{G}\mathbf{n} \tag{9}$$

where $\mathbf{n}$ is the process noise modeling target acceleration. This noise is assumed to be white with diagonal covariance, described by

$$\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}\delta(t-\tau)) \qquad \mathbf{D} = \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_a^2 \end{bmatrix} \tag{10}$$

where σ_a describes the typical acceleration of the target. The matrices $\mathbf{A}$ and $\mathbf{G}$ are given by

$$\mathbf{A} = \begin{bmatrix} 00 & 1 & 0 & \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{G} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{11}$$

For real-world applications, these equations must be discretized resulting in the model

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{v}_k \qquad \mathbf{v}_k \sim \mathcal{N}(0, \mathbf{Q}) \tag{12}$$

where $\mathbf{F}$ is the discretized state transition matrix, $\mathbf{x}_k$ the state at timestep k and $\mathbf{v}_k$ the discretized process noise with covariance $\mathbf{Q}$. In the real world, target kinematics cannot usually be described accurately by only a single motion model, possibly necessitating the full VIMMJIPDA. A well-tuned CV model should however give an acceptable baseline performance that other methods can be benchmarked against.

4.2. Sensor models

In a heterogeneous sensor fusion system different sensors can have different measurement models, $p(\mathbf{z}_k|\mathbf{x}_k)$ relating the received detections to the state estimates. Both active sensors receive measurements in the form of polar or spherical coordinates. Discarding height information from the lidar allows us to describe both measurements using an identical model albeit with different measurement noises,

$$f_z^{r/l}(\mathbf{x}_k) = \begin{bmatrix} \sqrt{p_x^2 + p_y^2} \\ \arctan(p_y/p_x) \end{bmatrix} + \mathbf{w}_k^{r/l} \qquad \mathbf{w}_k^{r/l} \sim \mathcal{N}(0, \mathbf{R}^{r/l}) \tag{13}$$

where r and l indicate respectively the radar or lidar. Due to non-linearities in the measurement function, this would usually require an extended Kalman filter (EKF). However, the measurements are converted back to Cartesian coordinates resulting in the measurement model

$$f_z^{r/l}(\mathbf{x}_k) = \begin{bmatrix} p_x \\ p_y \end{bmatrix} + \mathbf{w}_k^{r/l} \qquad \mathbf{w}_k^{r/l} \sim \mathcal{N}(0, \mathbf{J}\mathbf{R}^{r/l}\mathbf{J}^{\mathrm{T}}) \tag{14}$$

where $\mathbf{J}$ is the Jacobian of the polar to Cartesian conversion of the measurement and $\mathbf{R}$ is the measurement noise in polar coordinates. This eliminates the need for EKFs in the tracking system but requires an implementation with measurement-specific covariances.

For the passive camera sensor, the measurements begin as a series of pixel coordinates. If only bearing measurements are used we utilize the measurement model

$$f_z^{eo/ir}(\mathbf{x}_k) = \arctan(p_y/p_x) + \mathbf{w}_k^{eo/ir} \qquad \mathbf{w}_k^{eo/ir} \sim \mathcal{N}(0, \mathbf{R}^{eo/ir}) \tag{15}$$

where $\mathbf{R}$ is the sensor noise in bearing and *eo/ir* indicate the electro-optical or infrared cameras. If georeferencing (Helgesen et al., 2020) is applied to estimate detection ranges as well the measurement function is given by

$$f_z^{eo/ir}(\mathbf{x}_k) = \begin{bmatrix} p_x \\ p_y \end{bmatrix} + \mathbf{w}_k^{eo/ir} \qquad \mathbf{w}_k \sim \mathcal{N}(0, \mathbf{J}\mathbf{R}^{eo/ir}\mathbf{J}^{\mathrm{T}}) \tag{16}$$

where $\mathbf{R}$ is the measurement noise in pixel coordinates and $\mathbf{J}$ is the Jacobian of the pixel-to-Cartesian conversion for the measurement which is evaluated numerically in Helgesen et al. (2020). This increases the complexity of the measurement model but allows for a more accurate description of sensor uncertainty. Close to the camera, a single pixel will only cover a small area, further away this area increases. Describing uncertainty in pixel coordinates and then converting to Cartesian will compensate for this, increasing Cartesian uncertainty with the target range.

4.3. Association probabilities

Other key changes are made to the calculation of the association probabilities for the different hypotheses in the JIPDA. Different sensors have different performance characteristics, detection ranges are typically limited and detection performance depends on the range of the target. By introducing sensor-specific (s) parameters for the measurement noise, $\mathbf{R}^s$, and sensor and range (r) specific detection probabilities, $P_D^s(r)$ and false alarm intensities, $\lambda^s(r)$, the tracker is able to account for the differing characteristics of the various sensors across all ranges. Defining an association hypothesis, $\mathbf{a}_k$, as $\mathbf{a}_k = [a_k^1, \dots, a_k^n]$ where

$$a_k^t = \begin{cases} j & \text{if target } t \text{ claims measurement } j \\ 0 & \text{if target } t \text{ claims no measurements,} \end{cases} \tag{17}$$

the association probabilities are then given by

$$\Pr\{\mathbf{a}_k|\mathbf{Z}_{1:k}\} \propto \prod_{j \notin \mathbf{a}_k} \lambda^s(r_j) \prod_{t:a_k^t=0} (1 - \epsilon_{k|k-1}^t P_{\mathrm{D}}^s(r_{k|k-1}^t)\eta_{k|k-1}^t) \prod_{t:a_k^t>0} \epsilon_{k|k-1}^t P_{\mathrm{D}}^s(r_{k|k-1}^t)\eta_{k|k-1}^t l^{t,a_k^t} \tag{18}$$

where $\epsilon_{k|k-1}^t$ and $\eta_{k|k-1}^t$ are the predicted existence and observability probabilities of target t and $r_{k|k-1}^t$ the predicted target range. The number of false alarms is modeled as a Poisson point process where $\lambda^s(r_j)$ is the associated range-dependent intensity for sensor s and measurement j. For passive sensors without range information, a constant false alarm intensity is used. An important assumption in this model is that the number of false alarms in bounded sub-regions is completely independent of each other, i.e. the process is completely random. Compared to the rest of the elements in this product the distributions of the detection probability and false alarm intensities are assumed to be near-constant around the relevant track and measurement, we, therefore, approximate their distribution using mean values. State-dependent detection probability has also been explored in Musicki and Hanselmann (2008) using Gaussian mixtures and is also present in Williams (2015) along with non-homogeneous false alarm intensities. The track to measurement likelihood, l^{t,a_k^t}, is given by

$$\begin{aligned} l^{t,a_k^t} &= \int f_z^s(\mathbf{z}_k^{a_k^t}|\mathbf{x}_k^t) p_{k|k-1}^t(\mathbf{x}_k^t) \mathrm{d}\mathbf{x}_k^t \\ &= \mathcal{N}(\mathbf{z}_k^{a_k^t}; f_z^s(\hat{\mathbf{x}}_{k|k-1}^t), \mathbf{H}_k^{s,t}\mathbf{P}_{k|k-1}^t\mathbf{H}_k^{s,t\mathrm{T}} + \mathbf{R}^s) \end{aligned} \tag{19}$$

where $f_z^s(\mathbf{z}_k^{a_k^t}|\mathbf{x}_k^t)$ is the likelihood relating the measurement given by the association a_k^t to the target t and $p_{k|k-1}^t(\mathbf{x}_k^t)$ the predicted state distribution of the target. $\hat{\mathbf{x}}_{k|k-1}^t$ is the predicted state estimate while $\mathbf{P}_{k|k-1}^t$ is the corresponding state covariance. $\mathbf{H}_k^{s,t}$ is the measurement function Jacobian for sensor s and track t. These equations are based on the single-sensor derivation of the VIMMJIPDA found in Brekke et al. (2020).

4.4. Track management

Real-world applications of tracking in dynamic environments require some form of track management to initialize new tracks and terminate existing tracks. Newly formed tracks are given the label tentative tracks but are otherwise considered as fully confirmed tracks by the tracking system.

4.4.1. Track initialization

In any scan, the tracker performs data association between tracks, both tentative and confirmed, and the received measurements using a validation gate around the track. Once data association is complete any unassociated measurements, that is measurements not falling within the validation gate of an existing track, are used to form new tentative tracks. For each timestep, the current unassociated measurements are

compared to the unassociated measurements from the previous step from the same sensor based on the distance criterion

$$\|\mathbf{z}_k^i - \mathbf{z}_{k-1}^j\| \leq \mathbf{v}_{max} \Delta T_k \tag{20}$$

where $\mathbf{v}_{max}$ is the expected maximum speed of a target and ΔT_k is the time difference between the steps. If the criterion is met for a pair of measurements, a new tentative track is initialized. Tentative tracks are labeled confirmed if the test

$$\epsilon_k^t \geq T_{conf} \tag{21}$$

is met where T_{conf} is a tuning parameter. This test is performed on each preliminary track after an update and if passed the track is labeled as confirmed until its eventual termination. Both track types are treated identically by the tracking system and are thus considered jointly for data association, predictions, etc. The rest of the autonomy system could however use track status in their decision making, e.g. only considering confirmed tracks for path planning. Only active sensors are used in track initialization due to the lack of range information from the passive sensors.

4.4.2. Track termination

Existence-based approaches to tracking such as the JIPDA provide existence probability estimates as part of the tracking process. Once the update step is completed the existence probability of all tracks is evaluated using the existence criterion

$$\epsilon_k^t \geq T_r \tag{22}$$

where T_r is a threshold value. For each step without any measurement associations, a track will experience a reduction in existence probability. If the existence probability of a track drops low enough to fail this test the track is terminated and removed from the list of active tracks.

5. Performance measures

Accurate and reliable detection and tracking is an important part of a vessel's situational awareness. In this section, we present a series of performance measures used to evaluate both of these systems. All evaluations are performed automatically based on a data association process between the measurements and the true target position given by the ground truth, or the state estimate and the ground truth. In this process, tracks and measurements are assigned to their closest target given that the measurement error is less than a specified threshold. For the evaluation, we use a distance threshold of 20 m for the active sensors in environment 1 due to the size of the target Gunnerus. In environment 2 this threshold is 10 m. For the passive sensors, we use a bearing threshold of 10° in the detection evaluation. The Stone Soup Python framework (Thomas et al., 2017) or the MATLAB Sensor Fusion Toolbox which was used as a base in this work provide many of the measures presented here.

5.1. RMS error

Perhaps the most basic performance measure, the root mean square error (RMSE) yields information about the mean error of the individual sensors or tracks. Defining $\bar{\mathbf{e}}_i$ as the error, RMSE is calculated according to

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} |\bar{\mathbf{e}}_i|^2}. \tag{23}$$

For detection errors we first convert active sensor detections to Cartesian coordinates, yielding the position RMSE. For passive sensors, the bearing RMSE is used. For tracking evaluation, the position RMSE is used to evaluate the tracking estimates against the ground truth positions.

5.1.1. Measurement noise

Measurement noise plays an important part in the tuning of tracking systems. Inaccurate detection systems will have a higher degree of detection noise compared to more accurate systems, resulting in predictions being weighted more in the state estimation of targets. Detection noise is reported as the measurement error covariance matrix, given by

$$\mathbf{R} = \frac{1}{n-1}\sum_{i=1}^{n} (\mathbf{z}_i - f_z(\mathbf{x}_j))(\mathbf{z}_i - f_z(\mathbf{x}_j))^T \tag{24}$$

where f_z is the measurement function of the sensor.

5.2. Detection probability

Another important performance measure for a detection system is the probability that a target will be detected in any given scan. Several tracking methods also use this parameter in their bookkeeping to update a target's existence and detectability probability.

For a given sensor scan we assume that any target assigned to a measurement is detected. Keeping track of every targets assignment status allows the calculation of the detection probability, P_D, according to

$$P_D = \sum_{i=1}^{n} \frac{n_{det}^i}{n_{total}^i} \tag{25}$$

where n_{det} is the number of targets detected in scan i and n_{total} the number of targets present. Additional info can also be gained by examining the relationship between the detection probability of a target and the target range. We therefore report P_D in range-specific bins.

5.3. Clutter intensity

Any non-perfect sensor will eventually yield non-target measurements known as clutter or false alarms. The expected clutter intensity is the expected amount of false detections per sensor scan per area and it is an important tuning parameter in the JIPDA tracker. We assume that any measurements not associated with a target are clutter, yielding the calculation

$$\lambda = \frac{1}{\pi r_{max}^2} \frac{\sum_{k=1}^{n} m_k^{free}}{n_k} \tag{26}$$

where r_{max} is the maximum range of the sensor, n_k the total number of time steps, and m_k^{free} the number of unassociated measurements at time-step k. For the passive sensors the range term in the denominator is omitted. Readers should note that uniform false alarm rates in polar coordinates (e.g. radar) will not be uniform in Cartesian coordinates which could impact the tracking output when the expected number of false alarms is high. Due to the low false alarm rate of milliAmpere's sensors, we approximate this as uniform in Cartesian coordinates as well. False alarm intensities are evaluated both as uniform across the entire sensor range and in range-specific bins for the active sensors.

5.4. ANEES

The normalized estimation error squared (NEES) is a measure of statistical consistency. The tracking output can be said to be statistically consistent if the state errors are accurately described. NEES evaluates the relationship between the state error and the estimate covariance matrix according to

$$\text{NEES}_k = \bar{\mathbf{x}}_i \mathbf{P}_i^{-1} \bar{\mathbf{x}}_i \tag{27}$$

Taking the average of this across all time-steps yields the Average NEES (ANEES), which is chi-squared distributed. We report both the total ANEES for all states as well as the position-only ANEES.

$$\text{ANEES} = \sum_{k=1}^{n} \text{NEES}_k \tag{28}$$

This allows us to calculate a confidence interval for the ANEES.

5.5. Establishment length

Rapid track establishment is a desirable property allowing for increased safety when operating in environments where targets can appear close to the vessel. Track establishment occurs once a target is first associated with a track. We report the mean track establishment time from scenario start across all datasets.

5.6. Track breaks

This performance measure evaluates the tracking system's ability to keep tracks alive for valid targets. Certain situations, such as obscurement by larger vessels, might result in several sensor scans where the track receives no measurement associations. Track management could then be led to believe the track is false and remove it, resulting in a track break. We report both the total number of track breaks occurring as well as the length of the track breaks.

5.7. False tracks

False tracks are tracks that do not originate from valid targets but from false alarms. These are undesirable for several reasons, they increase the computational complexity of the tracking problem, they could interfere in the path-planning and collision avoidance parts of an autonomy system and they might prevent valid tracks from forming in their neighborhood. We report both the total number of false tracks as well as their cumulative length.

5.8. GOSPA

The General Optimal Subpattern Assignment (GOSPA) is a metric that evaluates multiple aspects of the tracking output taking into account localization errors, false tracks and missed targets. Defining the set of tracks at time k as $\mathbf{X}_k = [\mathbf{x}_k^1, \dots \mathbf{x}_k^m]$ and the set of truths as $\mathbf{Y}_k = [\mathbf{y}_k^1, \dots \mathbf{y}_k^n]$ the GOSPA metric is given by

$$\text{GOSPA} = \left(\min_{\pi \in \prod_n} \sum_{i=1}^{m} d^{(c)}(\mathbf{x}_k^i, \mathbf{y}_k^{\pi(i)})^p + \frac{c^p}{\alpha}(n-m) \right)^{\frac{1}{p}} \tag{29}$$

where $\prod_n$ is the set of all permutations of $\{1, \dots, n\}$. $d(x, y)$ is a metric for track-truth distance and $d^{(c)}(x, y) = \min(d(x, y), c)$ the distance cut-off given by the parameter c. In this work we use the unlabeled GOSPA with a Cartesian distance cut-off parameter of 20 m to match the track-truth assignment threshold used in other metrics. The rest of the parameters are set to $\alpha = p = 2$.

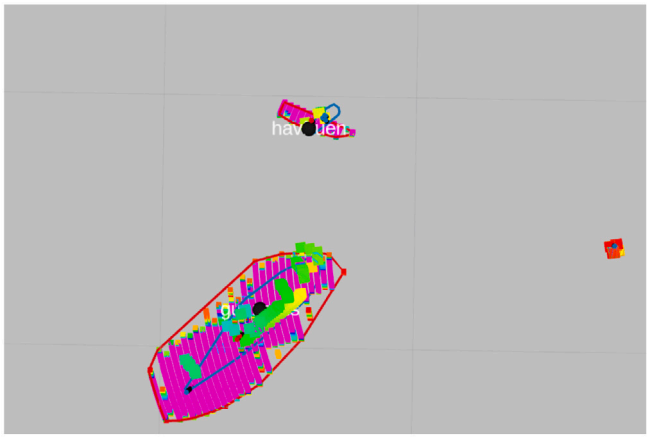
6. Datasets

Large amounts of sensor data were recorded over two days in two distinct maritime environments using the sensor systems onboard *milliAmpere*. In each environment, reference targets performed various maneuvers designed to test both the sensing and detection system as well as the tracking system in scenarios including challenges such as merged measurements, Fig. 6, and target obscuration, Fig. 5.

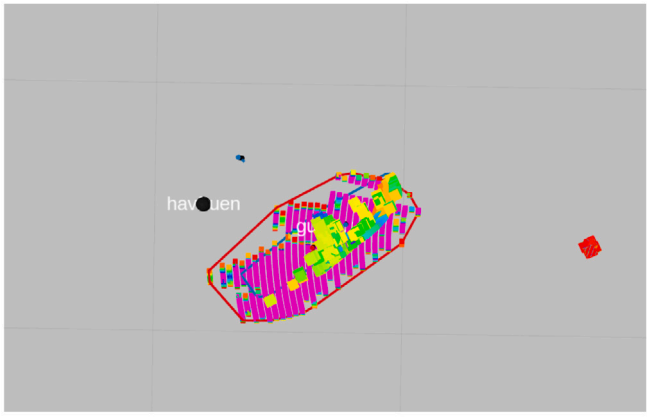
6.1. Environments

Two distinct environments were used in the data recording. Environment 1 consists of a large area outside Brattørkaia in Trondheim, Norway. This environment allows for large distances between the targets and the ASV and is similar to the coastal environments larger autonomous surface vehicles are expected to operate in. Reference targets can also operate at greater speeds and perform longer maneuvers compared to environment 2. These ranges challenge the long-range performance of the sensing system and might necessitate a greater reliance on the radar due to a more limited sensing range from the other sensors.

Environment 2 is situated in the channel between Brattøra and the city center in Trondheim. This is an urban environment lined with buildings and docks for leisure crafts providing returns for both the active and passive sensors. As the position of these features can be dynamic the land maps used to process active sensor data might not cover them resulting in a greater chance of false alarms. Processing of passive sensor detections to remove docked boats is not performed due



(a) Positive radar detection on target *Havfruen* (top cluster) before obscuration.



(b) Obscuration

Fig. 5. Negative detection of *Havfruen* during obscuration behind target *Gunnerus*. Radar point cloud is given by pink squares, lidar point cloud by colored squares. Blue lines are used for lidar cluster visualization, red lines for radar clusters. Ownship location is given by red square to the right and target ground truth by the black circles.

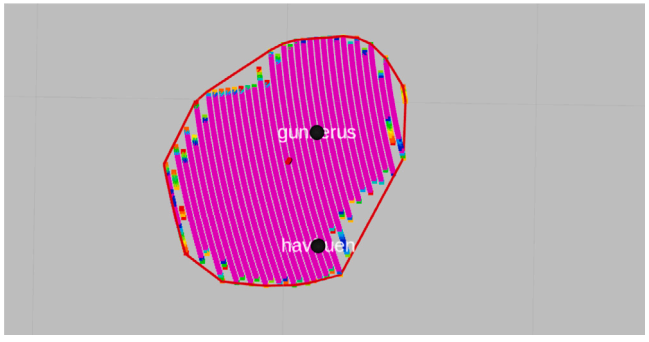


Fig. 6. Radar point cloud showing merged measurements. Target range to *milliAmpere* is 150 m.

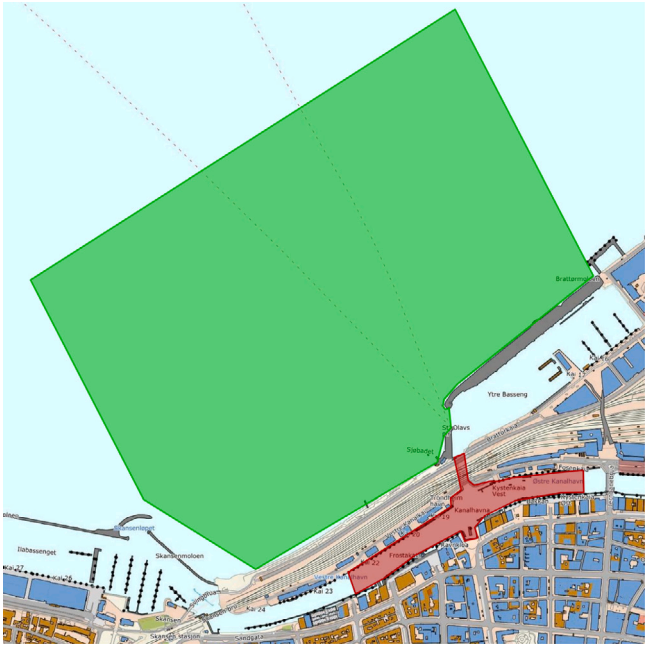


Fig. 7. Experimental area. Green polygon shows environment 1, red polygon environment 2. Map data is provided by the Norwegian Mapping Authority (©Kartverket).

Table 1
Target information.

	Havfruen	*Gunnerus*	*Jetboat*
Length	7.3 m	31.3 m	7.4 m
Width	2.8 m	9.6 m	2.9 m
GNSS	Ellipse-D	ZED-F9P as done in Vasstein (2021)	

to a lack of range information. Both speed and maneuverability are reduced compared to environment 1, however, targets can appear at much shorter ranges and typically operate at much closer distances introducing different challenges for ASVs. Both environments are shown in Fig. 7.

6.2. Reference targets

A total of three reference targets were employed during the collection of this dataset. Two targets were only used in a single environment each while one was used in both. All reference targets were equipped with 1 pair of high-precision Global Navigation Satellite System (GNSS) receivers as shown in Table 1. *Gunnerus* and *Jetboat* were equipped with GNSS receivers from Vasstein (2021). We later used post-processed kinematics from the open source library (RTKLIB, 2020) to refine precision and accuracy in this data. *Havfruen* was equipped with an SBG Systems Ellipse-D GNSS receiver.

6.2.1. Havfruen

Havfruen is a medium-sized leisure craft owned by the *Mannhullet* student association at NTNU. Due to its size and engine power, *Havfruen* is capable of rapid maneuvers at high speed putting greater emphasis on the motion model selection and tuning in the tracking system. *Havfruen* was used in both environments and is shown in Fig. 8.

6.2.2. Gunnerus

R/V *Gunnerus* is an ocean-going research vessel owned by NTNU and was used as a reference target in environment 1. Compared to *Havfruen*, *Gunnerus* is both slower and less maneuverable due to its far greater size (cf. Table 1), it is however more representative of vessels a larger

Fig. 8. *Havfruen* in environment 2.

Fig. 9. *Gunnerus* with *Havfruen* in the background in environment 1.

ASV might encounter while underways. Due to its size, *Gunnerus* casts a large sensor shadow area where other targets will not be detected as *Gunnerus* is obscuring them. This is particularly challenging for smaller targets operating at high speeds. Depending on the position of the ASV and the targets the ASV might only have seconds to detect and react to a possible collision putting a greater emphasis on rapid track establishment. *Gunnerus* is shown in Fig. 9.

6.2.3. Jetboat

The final reference target used in environment 2 is a water jetboat designed for professional use. Its low cross-sectional area challenges the detection system and could allow it to hide behind larger objects and docks in the channel at range. This target is shown in Fig. 10.

6.3. Environment 1

Data recording in this environment took place on the 4th of May 2021 from 09:44 to 13:41 in partially cloudy weather. 12 scenarios were recorded, however insufficient ground truth coverage meant several datasets had to be discarded, see Table 2 for an overview of ground truth coverage.

6.3.1. Scenario 2

In this scenario, *Havfruen* starts to the east while *Gunnerus* starts to the west. The targets perform a crossing maneuver in front of *milliAmpere* where *Havfruen* travels behind *Gunnerus* relative to *milliAmpere*. Challenges in this dataset include track merging, merged measurements, track continuation as well as target obscuration. A visualization is shown in Fig. 11.

Fig. 10. *Jetboat* in environment 2.

Table 2
Available data from GNNS receivers for environment 1. Red: not available, green: available.

GNNS ground truth coverage						
Scenario	*milliAmpere*		*Havfruen*		*Gunnerus*	
	Left	Right	Front	Back	Front	Back
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						

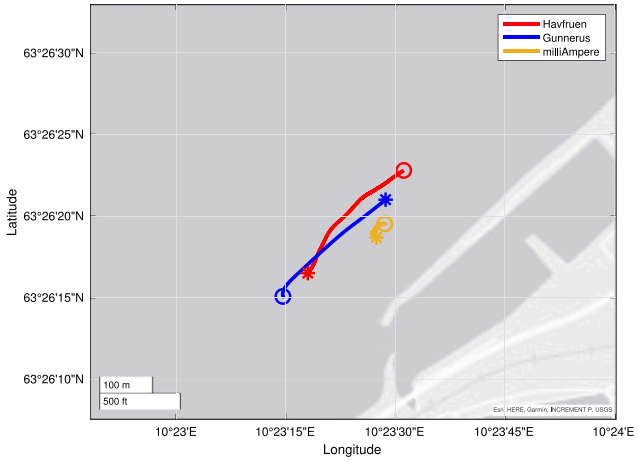


Fig. 11. Scenario 2. Circles signify starting positions, stars end positions.

6.3.2. Scenario 3

In scenario 3 *Havfruen* start to the north of *Gunnerus*. *Gunnerus* holds a relatively steady course southwest while *Havfruen* maneuvers around *Gunnerus*, passing to the south. Parts of this maneuver take place in the sensor shadow of *Gunnerus*, challenging the ability of the tracking system to both predict and keep alive obscured targets. A visualization is shown in Fig. 12.

6.3.3. Scenario 4

Both targets start west of *milliAmpere* traveling towards it side by side. Close to *milliAmpere* *Gunnerus* performs a maneuver to the north

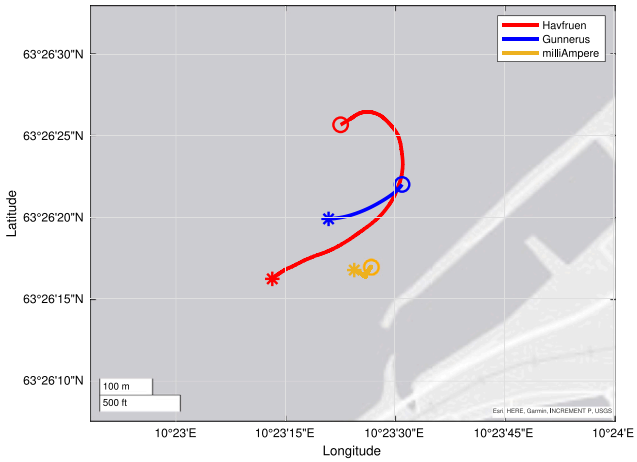


Fig. 12. Scenario 3. Circles signify starting positions, stars end positions.

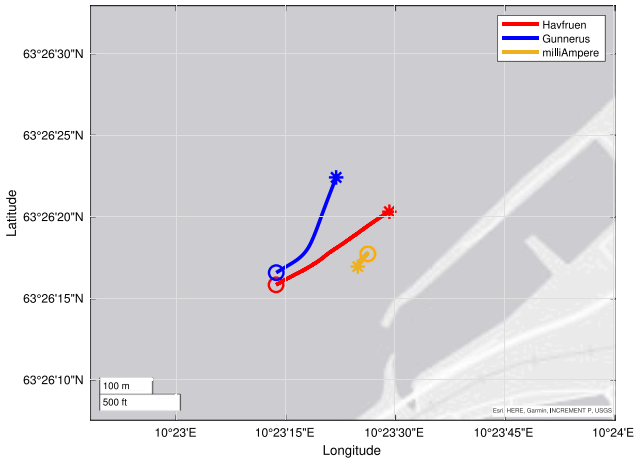


Fig. 13. Scenario 4. Circles signify starting positions, stars end positions.

while *Havfruen* continues on its original course. Challenges in this scenario include target maneuvers and merged radar measurements. A visualization is shown in Fig. 13.

6.3.4. Scenario 5

This scenario includes a high-speed overtake from *Havfruen* in the sensor shadow of *Gunnerus*. Both targets start north of *milliAmpere* to the east with *Havfruen* behind *Gunnerus*. Obscuration and track management are the main challenges in this dataset. A visualization is shown in Fig. 14.

6.3.5. Scenario 6

Havfruen starts behind *Gunnerus* with both targets situated westward of *milliAmpere*. *Gunnerus* and *Havfruen* both travel towards *milliAmpere* with *Havfruen* directly behind *Gunnerus* in its sensor shadow. Closer to *milliAmpere* *Gunnerus* maneuvers northwards revealing *Havfruen*. Challenges include obscuration and track establishment. A visualization is shown in Fig. 15.

6.4. Environment 2

Data recording in this environment took place on the following day, the 5th of May 2021 from 10:52 to 11:57 in partially cloudy weather. Similar ground truth persisted in this environment as well, see Table 3 for a complete overview.

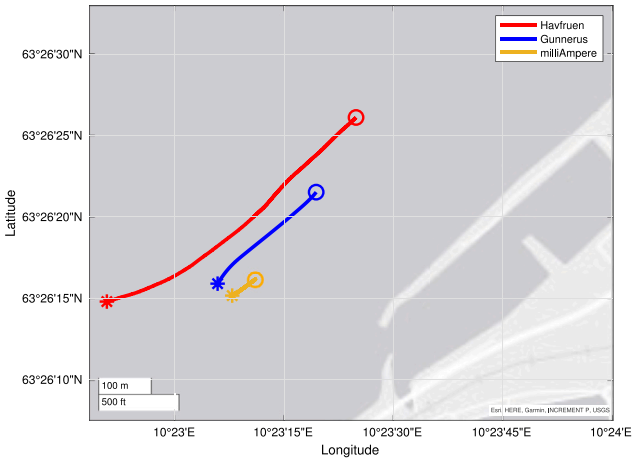


Fig. 14. Scenario 5. Circles signify starting positions, stars end positions.

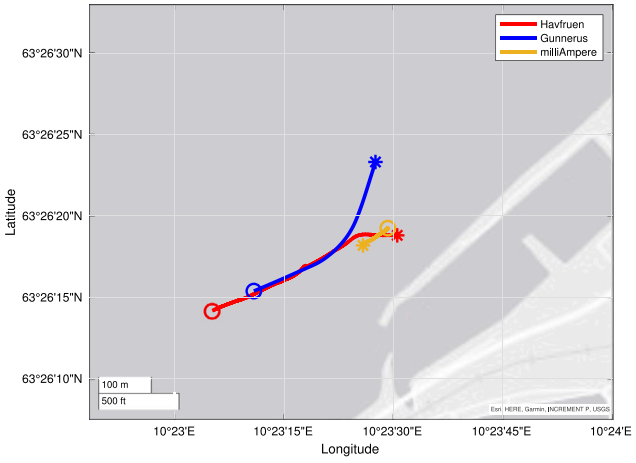


Fig. 15. Scenario 6. Circles signify starting positions, stars end positions.

Table 3
Available data from GNNS receivers for environment 2. Red: not available, green: available.

Scenario	GNNS ground truth coverage					
	milliAmpere		*Havfruen*		Jetboat	
	Left	Right	Front	Back	Front	Back
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						

6.4.1. Scenario 13

This scenario represents a common scenario in the channel with targets approaching *milliAmpere* from each side. *milliAmpere* is docked to the south which initially obscures both targets due to the surrounding buildings. Rapid track establishment is a key part of this scenario. A visualization is shown in Fig. 16.

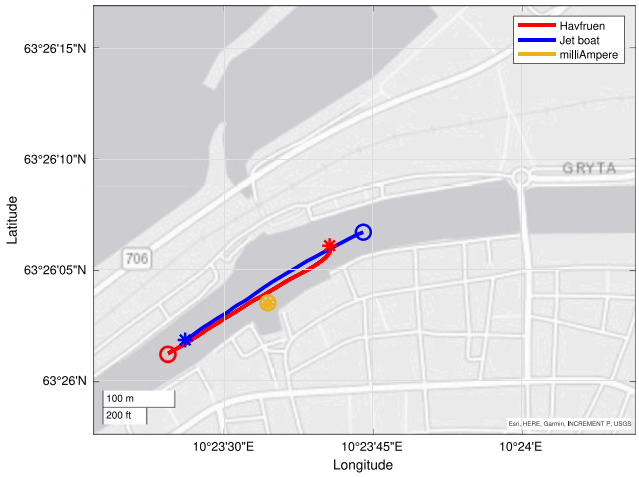


Fig. 16. Scenario 13. Circles signify starting positions, stars end positions.

Fig. 17. Scenario 16. Circles signify starting positions, stars end positions.

6.4.2. Scenario 16

In this scenario, we repeat the situation from scenario 13 but with *milliAmpere* stationary in the middle of the channel. This removes most of the obscurations from scenario 13 maximizing the detection range of the sensing system. Both targets travel directly towards *milliAmpere* and eventually maneuver around it. A visualization is shown in Fig. 17.

6.4.3. Scenario 17

Both targets approach *milliAmpere* from the east traveling side by side. This introduces the potential for merged measurements in the radar which could be mitigated by introducing additional sensors. A visualization is shown in Fig. 18.

6.4.4. Scenario 22

This scenario includes a channel crossing from *milliAmpere* starting north. Both targets start east and travel westward side by side. Near *milliAmpere* one target breaks off to the north continuing under the bridge. Track merging and track continuation are relevant challenges in this scenario as well as accurate tracking of target maneuvers. A visualization is shown in Fig. 19.

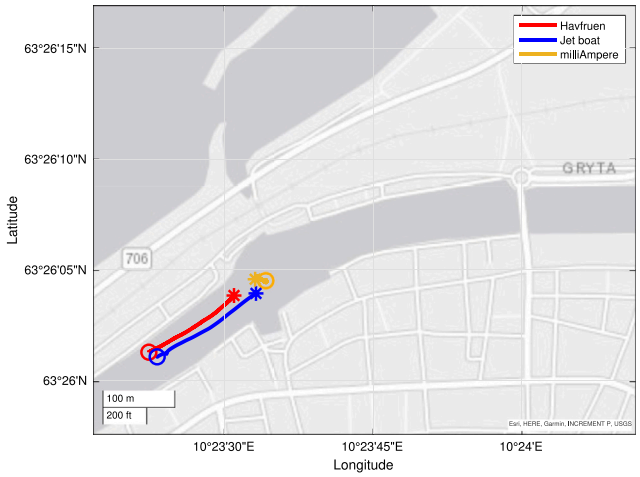


Fig. 18. Scenario 17. Circles signify starting positions, stars end positions.

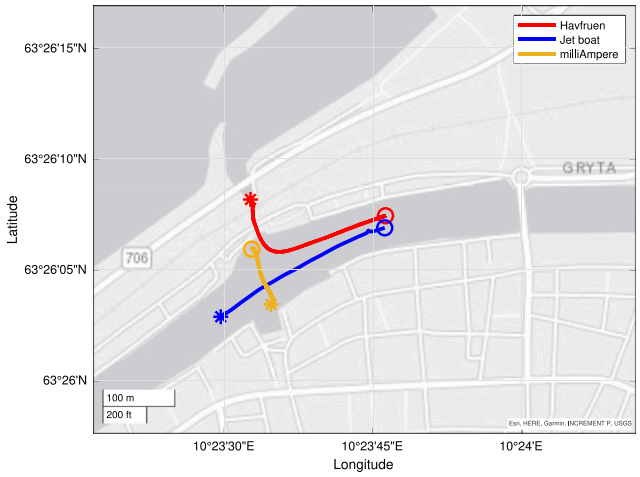


Fig. 19. Scenario 22. Circles signify starting positions, stars end positions.

7. Results

Based on the automatic evaluation system and performance measures presented in Section 5 we evaluate the performance of the sensing and tracking system based on the presented datasets. We evaluate scenarios from each environment separately to highlight the effect the surrounding area has on performance.

7.1. Sensing system

In terms of detection probability (P_D, cf. Fig. 20), we observe that the radar is supreme in environment 1 as expected. In more open coastal areas the radar has few challenges, especially for larger targets such as those used here. In more adverse conditions rain and sea clutter might require higher degrees of filtering, potentially removing weaker radar returns from smaller or less visible targets, however, this was not required for these datasets. Relative to the other sensors the radar is equal or leading across all ranges in detection probability, providing reliable detection out to a range of 300 m. After this detections are still present but less reliable. The lidar has a near-uniform high detection probability within its range but drops to zero outside this. One interesting observation is that the detection range of the lidar exceeds the manufacturer specified max range possibly owing to the size of the targets.

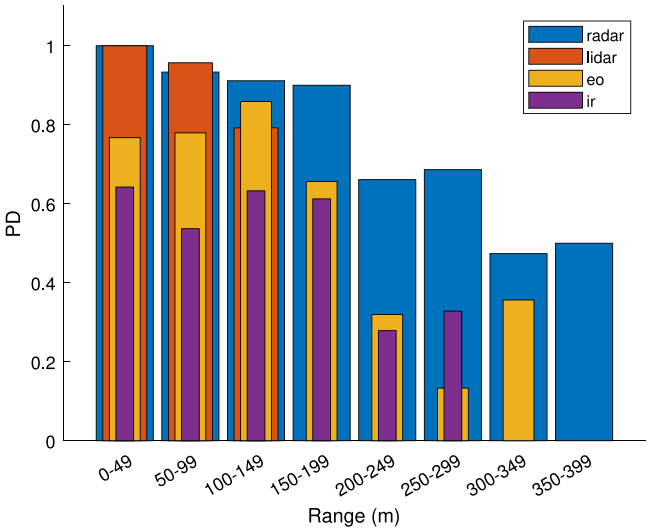


Fig. 20. Detection probability, environment 1.

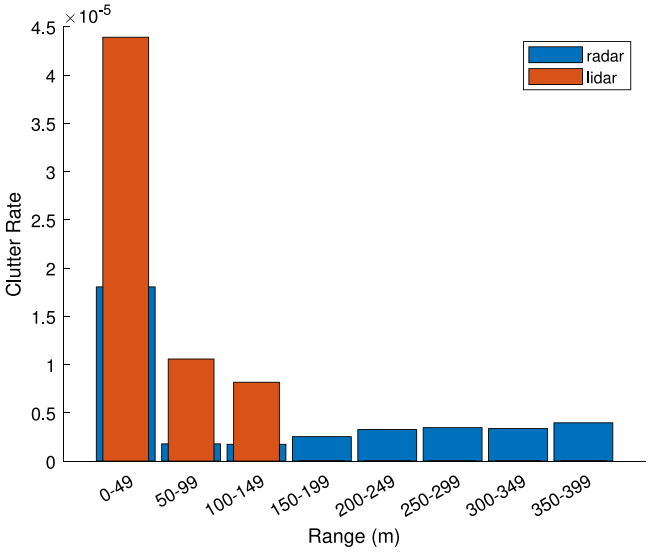


Fig. 21. Active sensor false alarm intensities, environment 1.

Both active sensors also show near-identical sensor noise in range and bearing (cf. Table 4). We observed some cases of merged measurements from the radar when the targets were moving close to each other. This might be explained by the relatively large beamwidth of the radar and would have an adverse effect on the bearing performance of the radar. False alarm intensity (cf. Fig. 21) is higher for the lidar which exhibits a tendency to generate false alarms from target wakes. In more challenging conditions with higher sea states and/or precipitation, this is expected to change. RMS position error is slightly lower for the radar but not significantly so.

The imaging sensors also exhibit similar behavior in terms of detection probability. For both sensors, it remains near-constant before dropping off sharply. At more extreme ranges we see that the detection probabilities converge, however, at every other range, the EO cameras perform better. Similar error dynamics are also observed for both cameras. The sensor noise is near identical for both sensors, the same is the RMS error. False detections are less of an issue with the infrared cameras but the false alarm intensity is still very equal.

In environment 2 (cf. Fig. 22), the radar detection probability shows some degradation compared to environment 1. Urban environments

Table 4
Detection performance, environment 1. **R** is the diagonal elements of the sensor noise matrix (covariance). Radar/Lidar noise is in polar coordinates, RMSE in Cartesian coordinates. EO/IR is defined in bearings. Detection probability is the non-zero average across all ranges.

	Radar	Lidar	EO	IR
R	$\begin{bmatrix}27.54\\0.0060\end{bmatrix}$	$\begin{bmatrix}33.78\\0.0054\end{bmatrix}$	0.0015	0.0018
RMSE	8.65	8.82	0.064	0.083
λ	4.13e−6	1.42e−5	0.1658	0.1583
P_D	0.76	0.92	0.51	0.55

Table 5
Detection performance, environment 2. **R** is the diagonal elements of the sensor noise matrix (covariance). Radar/Lidar noise is in polar coordinates, RMSE in Cartesian coordinates. EO/IR is defined in bearings. Detection probability is the non-zero average across all ranges.

	Radar	Lidar	EO	IR
R	$\begin{bmatrix}12.09\\0.0031\end{bmatrix}$	$\begin{bmatrix}3.79\\0.0024\end{bmatrix}$	0.0029	0.0025
RMSE	4.43	2.79	0.085	0.072
λ	1.89e−5	4.77e−6	2.1387	0.8229
P_D	0.70	0.75	0.78	0.32

are especially challenging for the radar due to the effects of multipath propagation and clutter returns from buildings and objects (Tahmoush et al., 2012). These can also obscure the targets rendering them unobservable for the sensor. Lidar performance is also degraded at longer ranges, however, this might be explained by the reduction in target size from environment 1 to environment 2. At closer ranges (0-75 m) performance is near identical.

In addition to detection probability, we also observe performance changes in terms of detection accuracy in the urban environment. Sensor noise covariance is significantly reduced for both active sensors and we also observe a reduction in the RMS position error of the measurements, Table 5. For the radar, this might be explained by a reduction in merged measurements. For the lidar, the reduction in target size could have contributed positively as long-range detections of targets heading towards or away from *milliAmpere* tend to only contain a few points along the front or back of the vessel yielding large position errors for targets like *Gunnerus*. False alarm intensity is still low and while the lidar outperforms the radar overall we observe that the two sensors are very closely matched (cf. Fig. 23). Aggressive land filtering was applied to achieve this result. However, due to the dynamic nature of the environment land maps might require frequent updates or a more SLAM-like solution with dynamic map updates if these results are to be maintained.

At shorter ranges, up to 100 m, the detection probability of the EO cameras is somewhat improved relative to the coastal environment. Challenging lighting conditions in environment 1 might explain some of this improvement. Environment 1 had partial cloud coverage with intermittent sunshine and the dynamic range of the lighting was therefore quite large. Underexposure was experienced in some images which might have impacted the detection probability. At ranges exceeding 100 m detection probability degrades in the urban environment, possibly due to increased background complexity in the images or reduced target sizes. Infrared performance degraded significantly, in environment 1 the detection probability hovered around 0.6 up to 200 m range. In environment 2 detection probability is higher at the very shortest ranges but rapidly degrades to below 0.5 at ranges greater than 50 m. Some of this might be corrected with further training, however manual inspection of the infrared images revealed that separating a valid target from the background was a significant challenge at longer ranges.

In contrast with the active sensors, we observe increased sensor noise for both imaging sensors. Compared to environment 1, the infrared cameras now exhibit lower sensor noise than the EO cameras as well as a false alarm intensity one order of magnitude smaller. Combined with the low detection probability it is likely that both are caused by a significant reduction in detections, both valid and false alarms, from the IR pipeline. Readers should be aware that many of the camera measurements classified as false alarms originate from the many docked vessels in environment 2. The active sensor pipelines filter these out using extended land maps, the same methods cannot be applied to the bearing detections of the cameras due to a lack of range information.

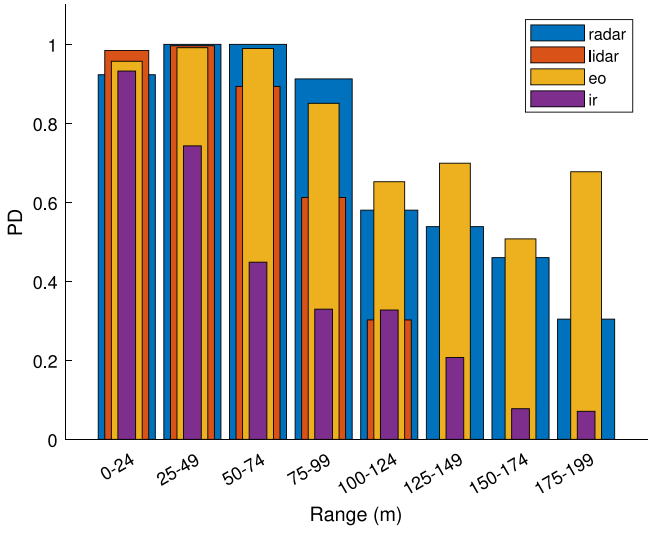


Fig. 22. Detection probability, environment 2.

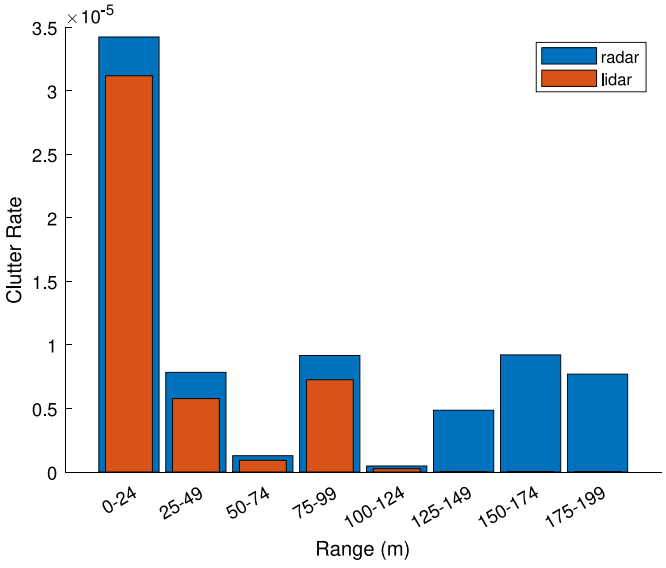


Fig. 23. Active sensor false alarm intensities, environment 2.

7.2. Tracking system

While detection performance plays a key part in the performance of a situational awareness system it is the tracking process that yields the actual state estimates of other targets. In this section, we present the tracking performance of a selection of sensor combinations for each individual environment.

7.2.1. Tuning parameters

The detection performance evaluation provides many of the required tuning parameters of the tracking system with the notable exception of the process noise of the target motion model. This parameter was tuned on scenario 2 using radar measurements against the ANEES performance measure. Some changes were required in the camera sensor noise covariances to improve statistical consistency. The rest of the parameters were also tuned on this dataset to minimize track breaks. The final tuning parameters are presented in Appendix.

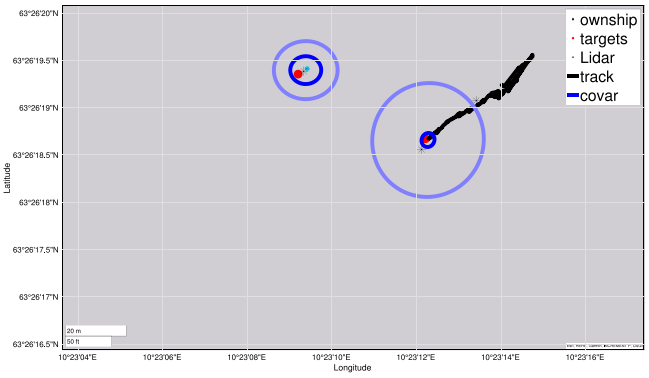
7.2.2. Environment 1

The detection performance of the various sensors is well reflected in the tracking performance as seen in Table 6. Based on the long-range detection performance of the radar versus the lidar one would expect the radar to establish tracks faster, as most targets are situated outside the lidar detection range when the various scenarios start. As expected the tracking results reveal this is true with the average lidar track establishment length being almost five times longer. Radar establishment lengths could be even shorter, the presence of merged measurements at the start of scenario 4, shown in Fig. 6, and target obscuration in scenario 6, drastically impacts the establishment lengths of the radar. Removing these scenarios from the evaluation reduces the establishment length of the radar to only 2.2 s which is equivalent to two scans. Position RMS errors are also close to the sensor evaluation with the lidar slightly trailing the radar for *Gunnerus*. Due to target obscuration in scenario 2 the *Havfruen* track coalesces on *Gunnerus*, inflating the RMS position error for this target for radar-only tracking. Reducing the process noise to 0.5 eliminates this and yields a position RMS error for *Havfruen* of only 10.80 m. This does however inflate ANEES, reducing statistical consistency. Track breaks are far shorter with the radar due to its long-range detection performance. Both sensors struggle slightly with underestimating the error covariance leading to reduced statistical consistency, however, the radar is closer to the expected value. In terms of GOSPA, the lidar has a slight lead which might be explained by fewer and shorter false tracks.

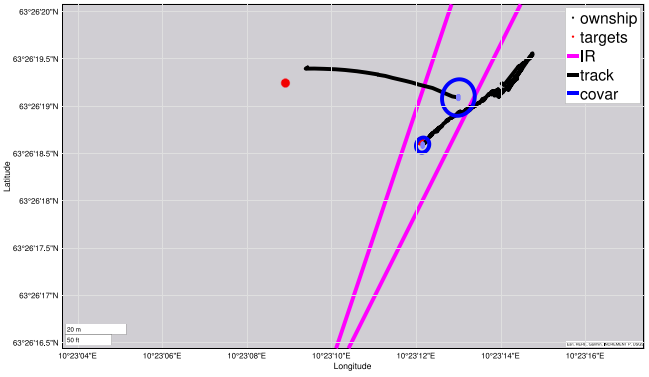
Introducing sensor fusion yields several performance benefits. Fusing radar and lidar improves RMS position error compared to radar only tracking for both targets. Compared to lidar only tracking *Havfruen* experiences a minor reduction in position accuracy while *Gunnerus* improves. Track break lengths are also marginally worse compared to radar-only tracking but still much better than lidar-only tracking. If different target vessels with less radar reflectivity such as kayaks are used, the effects of radar–lidar fusion are expected to provide greater performance benefits.

When fused with lidar the passive EO cameras introduce an improvement in track breaks as well as positioning accuracy at the cost of some statistical consistency. When fused with radar measurements we observe a degradation in position accuracy for *Havfruen* while *Gunnerus* remains similar, possibly due to *Gunnerus* having a larger detection probability due to its size. Track breaks are also significantly longer but still better than lidar-only tracking. Fusing both active sensors with one or more passive sensors does not produce any significant benefits across most performance measures. The introduction of the infrared cameras does however lead to a large increase in position error for the *Havfruen* target. The cause of this is a series of missed detections from *Havfruen* in both scenarios 2 and 5 which results in both tracks coalescing on *Gunnerus*, massively increasing the position error of the track associated with *Havfruen*. An illustration of this is shown in Fig. 24.

When it comes to false tracks the radar has a much larger problem than the lidar. While the difference in the number of false tracks is fairly low, 7 and 27, the length of the false tracks is 10 times longer for the radar. This suggests that the lidar false tracks are introduced by transient false alarms while the radar false tracks originate in measurements that are consistently present such as land or sea markings. Fusing multiple sensors does seem to provide a benefit in reducing the length of false tracks as false alarms in one type of sensor are unlikely to be present in other types of sensors. This would reduce the existence probability of any false track leading to earlier track death.



(a) Time: 63.6s



(b) Time: 64.7s

Fig. 24. Track coalescence on *Gunnerus*, scenario 5.

Table 6
Tracking performance, environment 1. Due to space constraints units have been excluded. Establishment lengths (Est.L), false track lengths (F.T.L) and break lengths (Break.L) are in seconds while position RMS error (posRMSE) is reported in meters per target. GOSPA is reported as RMS.

Sensors	Est.L	Break.L	posRMSE	GOSPA	ANEES
L	47.0	83.7	13.98;11.06	17.43	8.43
R	11.2	8.4	29.28;8.92	20.05	5.65
L,R	11.9	10.0	15.97;8.26	18.58	7.72
L,EO	47.6	63.4	10.57;10.96	16.81	11.04
R,EO	12.8	37.2	49.80;9.60	19.41	8.45
L,R,EO	12.1	32.8	85.58;9.91	18.65	9.82
L,R,IR,EO	12.2	86.3	38.91;9.43	20.37	15.84

Sensors	Num. false tracks	False track length
L	7	61.5
L,R	42	510.9
L,R,EO	75	414.7
L,R,IR,EO	73	402.5

7.2.3. Environment 2

In the urban environment (cf. Table 7), we observe several changes in the tracking performance for the various sensors. Both sensors have significantly reduced position errors, possibly due to reduced target sizes, and the radar also has better statistical consistency. Track break length for the radar has increased by an order of magnitude highlighting the challenges of radar tracking in urban environments. The lidar switches from underestimating the error covariance to overestimating it, an outcome that was expected based on the sensor performance evaluation. Track establishment lengths remain similar for both active sensors compared to environment 1.

False tracks are a much greater issue in this environment for the radar. While the number of false tracks remains similar their length

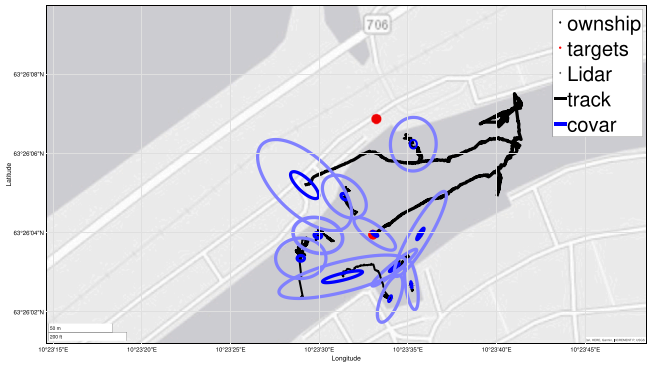


Fig. 25. Track divergence due to EO detections from docked boats.

is nearly doubled indicating that most tracks originate from persistent returns such as docked boats falling outside the land map used for filtering. This is also reflected in the GOSPA metric, despite the improvement in position errors GOSPA is worse compared to environment 1 due to track breaks and false tracks. Lidar false tracks remain similar across both environments leading to an improved GOSPA score due to reduced positioning errors.

Introducing sensor fusion yields some unexpected results. Lidar–radar fusion has a negligible effect on establishment length and position accuracy, however, track breaks are far longer than each individual sensor. This is most likely due to the tuning parameters originating from the sensor evaluation in environment 1 where the lidar had greater long-range performance due to target sizes. In environment 2 the tracker expects the lidar to perform similarly, reliably detecting targets out to ranges of 150 m. In reality, a sharp performance drop occurs after 100 m which can lead to premature track death. Fusing passive sensors with lidar does yield a moderate reduction in track breaks at the cost of some position accuracy, however, statistical consistency is improved. If radar is added tracking performance degrades significantly across many metrics. One potential explanation is that the high update frequency of the cameras compared to the radar allows detections from docked boats along the channel to capture valid tracks causing them to diverge from the actual target (cf. Fig. 25). As track-truth associations are only broken when the original track is either dead or a new track has been assigned to the truth, this would lead to a significant increase in position errors. Passive fusion also has a negative effect on false tracks. Due to the ambiguity of bearing measurements and the large number of false alarms most false tracks will experience increased measurement association thus keeping them alive for longer periods of time. This is also reflected in GOSPA which is higher for all combinations of active–passive fusion compared to active-only tracking.

8. Discussion and future work

The performance evaluation of this work points to several potential performance improvements. Detection performance is adequate for most sensors in terms of detection probability, however, improvements are possible for the IR cameras, especially in urban environments. Additional improvements could also be made to the handling of false detections for the imaging sensors. This is especially visible in environment 2 where tracking performance degrades significantly compared to environment 1. While this environment is towards the more extreme in terms of detections from docked boats it does well in highlighting the need for detection filtering for imaging sensors as well. Applying georeferencing (Helgesen et al., 2020) to estimate ranges is a possible solution, allowing the land filtering used for the active sensors to be applied to the cameras as well.

Stationary clutter measurements are not a passive sensor exclusive issue. In environment 1 the land returns from the radar will in certain

Table 7
Tracking performance, environment 2. Due to space constraints units have been excluded. Establishment lengths (Est.L) and break lengths (Break.L) are in seconds while position RMS error (posRMSE) is reported in meters per target. GOSPA is reported as RMS.

Sensors	Est.L	Break.L	posRMSE	GOSPA	ANEES
L	44.3	76.2	4.49;4.68	15.97	1.34
R	11.8	95.7	9.68;6.94	25.97	4.69
L.R	11.1	188.7	8.22;4.58	24.00	1.83
L,EO	56.2	45.3	20.7;14.75	17.45	4.76
R,EO	13.8	68.0	37.73;75.23	32.1	28.28
L,R,EO	17.4	111.3	30.44;29.93	31.01	12.96
L,R,IR,EO	23.8	200.2	83.53;50.49	30.97	51.90

Sensors	Num. False Tracks	False Track Length
L	11	47.4
R	28	1026.8
L,R	59	806.13
L,EO	8	117.11
R,EO	66	1417.7
L,R,EO	70	1591.5
L,R,IR,EO	131	2244.7

Table 8
Sensor tuning parameters. **R** is given as the diagonal elements of the covariance matrix, active sensors in polar coordinates and passive sensors as bearings. Active sensor false alarm intensities are given in the table below.

	R	λ
Lidar	[33.7763, 0.0054]	N/A
Radar	[27.5351, 0.0060]	N/A
EO	0.0203	0.17
IR	0.0288	0.16

Table 9
Range dependent false alarm intensities (λ).

	Lidar	Radar
0 m–49 m	4.39e−5	1.81e−5
50 m–99 m	1.06e−5	1.78e−6
100 m–149 m	8.16e−6	1.73e−6
150 m–199 m	0	2.53e−6
200 m–249 m	0	3.27e−6
250 m–299 m	0	3.46e−6
300 m–349 m	0	3.36e−6
349 m–399 m	0	3.96e−6

places extend into the water as seen in Fig. 4. Due to these measurements falling in areas with potential targets a trade-off has to be made between long-range false tracks and the detection of potential targets. At closer ranges, these measurements disappear possibly necessitating a dynamic land map or discarding measurements outside a certain range threshold. For more maneuverable ASVs this could be a valid option as most targets can be avoided at shorter notice.

In a tracking context, these measurements are also an issue as they violate the assumptions of the clutter model which assumes all clutter measurements are independent. Changing the clutter model to account for this is another possible solution, this might however require prior knowledge of the operating area to accurately model clutter densities. Another avenue of research is to introduce sensor-specific observability probabilities to complement the range-dependent detection probabilities.

Several improvements to how the track management system interacts with multiple sensors are also possible. With the current method of track initialization, a target requires two consecutive measurements from the same active sensor. Allowing tracks to be confirmed by multiple sensors could yield faster track establishment, especially if passive sensors are also included.

Recordings from the GNSS receivers (cf. Tables 4 and 5) proved to be a challenge due to several reasons. The manpower required to

Table 10
Range dependent detection probabilities.

	Lidar	Radar	EO	IR
0 m–49 m	0.99	0.99	0.77	0.64
50 m–99 m	0.96	0.93	0.77	0.54
100 m–149 m	0.79	0.91	0.86	0.63
150 m–199 m	0	0.90	0.66	0.61
200 m–249 m	0	0.66	0.32	0.28
250 m–299 m	0	0.68	0.13	0.33
300 m–349 m	0	0.47	0.36	0
349 m–399 m	0	0.5	0	0

Table 11
JIPDA tuning parameters.

σ_a	1.5	Process noise
g	3	Gate size
η_0	0.5	Init. exi. prob.
ϵ_0	1	Init. obs. prob.
T_{conf}	0.8	Track confirmation threshold
T_r	0.25	Track termination threshold
$\mathbf{F}_\eta$	$\begin{bmatrix} 0.99 & 0.01 \\ 0 & 1 \end{bmatrix}$	Exi. transition matrix
$\mathbf{F}_\epsilon$	$\begin{bmatrix} 0.9 & 0.1 \\ 0.48 & 0.52 \end{bmatrix}$	Obs. transition matrix
$\mathbf{v}_{max}$	10 m/s	Track init. max speed

simply conduct the experiments meant no personnel could be dedicated to monitoring the GNSS equipment during recordings. A mix of proprietary and open-source receivers made it impossible to create any common remote monitoring or error system to alert the participants of record failures. The use of redundant receivers on 29 scenarios did give decent scenario coverage in the end. It is suggested that future experiments should include a remotely operated system for monitoring the status of the individual receivers. This would lead to better scenario coverage, the possibility for estimating precise heading and give a basic sensor validation for the ground truth data.

9. Conclusion

In this work, we have presented an open dataset for maritime tracking with four heterogeneous sensors across two distinct environments and evaluated a complete detection and tracking system on this data. Several detection issues were uncovered, such as merged measurements and sensor shadowing. Both general and multi-sensor specific tracking issues were found such as stationary clutter measurements and differing detection ranges requiring range-specific detection probabilities. The performance evaluation of both the detection and tracking system also revealed several environment-specific issues that could be mitigated by introducing multiple types of sensors, either individually or in fusion, and that sensor fusion provided several performance benefits for tracking.

CRediT authorship contribution statement

Øystein Kaarstad Helgesen: Experiment planning and participation, Multi-sensor JIPDA adaption, Sensor fusion pipeline, Dataset processing, Writing. **Kjetil Vasstein:** Experiment planning and participation, Ground truth (collection, processing, and analysis), Writing. **Edmund Førland Brekke:** Experiment planning, Implementation of single-sensor JIPDA, Writing. **Annette Stahl:** Experiment planning, Writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the Research Council of Norway through Projects 223254, 331921 and 309230, and by NTNU through the Autoferry project. Bjørn-Olav Holtung Eriksen and Erik Wilthil provided the radar and lidar pipeline used in this work. Michael Ernesto Lopez piloted *milliAmpere* during the data collection. Audun Gullikstad Hem coordinated onboard the target ship Gunnerus.

Appendix. Tuning parameters

The various tuning parameters used in this work are shown in Tables 8, 9, 10 and 11 and are based on the sensor evaluation from environment 1.

References

Bar-Shalom, Y., Kirubarajan, T., Li, X.-R., 2002. Estimation with applications to tracking and navigation. John Wiley & Sons, Inc. New York, NY, USA.

Bar-Shalom, Y., Tse, E., 1975. Tracking in a cluttered environment with probabilistic data association. Automatica 11 (5), 451–460. http://dx.doi.org/10.1016/0005-1098(75)90021-7, URL: http://www.sciencedirect.com/science/article/pii/0005109875900217.

Bochkovskiy, A., Wang, C., Liao, H.M., 2020. YOLOv4: Optimal speed and accuracy of object detection. CoRR, arXiv:2004.10934.

Brekke, E.F., Hem, A.G., Tokle, L.-C.N., 2020. The VIMMJIPDA: Hybrid state formulation and verification on maritime radar benchmark data. In: Global Oceans 2020: Singapore – U.S. Gulf Coast. pp. 1–5. http://dx.doi.org/10.1109/IEEECONF38699.2020.9389007.

Brekke, E.F., Hem, A.G., Tokle, L.-C.N., 2021. Multitarget tracking with multiple models and visibility: Derivation and verification on maritime radar data. IEEE J. Ocean. Eng. 46 (4), 1272–1287. http://dx.doi.org/10.1109/JOE.2021.3081174.

Chavez-Garcia, R.O., Aycard, O., 2016. Multiple sensor fusion and classification for moving object detection and tracking. IEEE Trans. Intell. Transp. Syst. 17 (2), 525–534. http://dx.doi.org/10.1109/TITS.2015.2479925.

Cheng, Y., Jiang, M., Zhu, J., Liu, Y., 2021. Are we ready for unmanned surface vehicles in inland waterways? The USVInland multisensor dataset and benchmark. IEEE Robot. Autom. Lett. 6 (2), 3964–3970. http://dx.doi.org/10.1109/LRA.2021.3067271.

Cho, H., Seo, Y.-W., Kumar, B.V., Rajkumar, R.R., 2014. A multi-sensor fusion system for moving object detection and tracking in urban driving environments. In: 2014 IEEE International Conference on Robotics and Automation. ICRA, pp. 1836–1843. http://dx.doi.org/10.1109/ICRA.2014.6907100.

Cormack, D., Schlangen, I., Hopgood, J.R., Clark, D.E., 2020. Joint registration and fusion of an infrared camera and scanning radar in a maritime context. IEEE Trans. Aerosp. Electron. Syst. 56 (2), 1357–1369. http://dx.doi.org/10.1109/TAES.2019.2929974.

Fortmann, T.E., Bar-Shalom, Y., Scheffe, M., 1980. Multi-target tracking using joint probabilistic data association. In: 1980 19th IEEE Conference on Decision and Control Including the Symposium on Adaptive Processes. pp. 807–812. http://dx.doi.org/10.1109/CDC.1980.271915.

Fowdur, J.S., Baum, M., Heymann, F., 2021. Real-world marine radar datasets for evaluating target tracking methods. Sensors 21 (14), 4641. http://dx.doi.org/10.3390/s21144641.

Gaglione, D., Braca, P., Soldi, G., 2018. Belief propagation based AIS/radar data fusion for multi - target tracking. In: 2018 21st International Conference on Information Fusion. FUSION, pp. 2143–2150. http://dx.doi.org/10.23919/ICIF.2018.8455217.

Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 3354–3361. http://dx.doi.org/10.1109/CVPR.2012.6248074.

Haghbayan, M.-H., Farahnakian, F., Poikonen, J., Laurinen, M., Nevalainen, P., Plosila, J., Heikkonen, J., 2018. An efficient multi-sensor fusion approach for object detection in maritime environments. In: 2018 21st International Conference on Intelligent Transportation Systems. ITSC, pp. 2163–2170. http://dx.doi.org/10.1109/ITSC.2018.8569890.

Helgesen, O.K., Brekke, E.F., Helgesen, H.H., Engelhardtsen, Ø., 2019. Sensor combinations in heterogeneous multi-sensor fusion for maritime target tracking. In: 2019 22th International Conference on Information Fusion. FUSION, pp. 1–9.

Helgesen, O.K., Brekke, E.F., Stahl, A., Engelhardtsen, O., 2020. Low altitude georeferencing for imaging sensors in maritime tracking. IFAC-PapersOnLine 53 (2), 14476–14481. http://dx.doi.org/10.1016/j.ifacol.2020.12.1449, 21st IFAC World Congress. URL: https://www.sciencedirect.com/science/article/pii/S2405896320318607.

Kniaz, V.V., Knyaz, V.A., Hladůvka, J., Kropatsch, W.G., Mizginov, V.A., 2018. ThermalGAN: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset. In: Computer Vision – ECCV 2018 Workshops. Springer International Publishing.

Landsnes, K., 2021. Domain adaptation for detection of maritime vessels in images. Norwegian University of Science and Technology, URL: https://hdl.handle.net/11250/2831747.

Lin, T., Maire, M., Belongie, S.J., Bourdev, L.D., Girshick, R.B., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: Common objects in context. CoRR arXiv:1405.0312.

Musicki, D., Evans, R., 2004. Joint integrated probabilistic data association: JIPDA. IEEE Trans. Aerosp. Electron. Syst. 40 (3), 1093–1099.

Musicki, D., Evans, R., Stankovic, S., 1992. Integrated probabilistic data association (IPDA). In: Proceedings of the 31st IEEE Conference on Decision and Control. pp. 3796–3798 vol.4. http://dx.doi.org/10.1109/CDC.1992.370951.

Musicki, D., Hanselmann, T., 2008. State dependent detection and object tracking. In: 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems. pp. 372–377. http://dx.doi.org/10.1109/MFI.2008.4648094.

Özkanoğlu, M.A., Ozer, S., 2022. InfraGAN: A GAN architecture to transfer visible images to infrared domain. Pattern Recognit. Lett. http://dx.doi.org/10.1016/j.patrec.2022.01.026, URL: https://www.sciencedirect.com/science/article/pii/S0167865522000332.

Ren, S., He, K., Girshick, R.B., Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. CoRR, arXiv:1506.01497.

RTKLIB, 2020. RTKLIB: AN open source program package for GNSS positioning. version 2.4.3 b34, http://www.rtklib.com/.

Schöller, F., Blanke, M., Plenge-Feidenhans', M., Nalpantidis, L., 2020. Vision-based object tracking in marine environments using features from neural network detections. IFAC-PapersOnLine 53 (2), 14517–14523. http://dx.doi.org/10.1016/j.ifacol.2020.12.1455, 21st IFAC World Congress. URL: https://www.sciencedirect.com/science/article/pii/S240589632031867X.

Schuster, M., Blaich, M., Reuter, J., 2014. Collision avoidance for vessels using a low-cost radar sensor. IFAC Proc. Vol. 47 (3), 9673–9678. http://dx.doi.org/10.3182/20140824-6-ZA-1003.01872, 19th IFAC World Congress. URL: https://www.sciencedirect.com/science/article/pii/S1474667016431447.

Tahmoush, D., Silvious, J., Bender, B., 2012. Radar surveillance in urban environments. In: 2012 IEEE Radar Conference. pp. 0220–0225. http://dx.doi.org/10.1109/RADAR.2012.6212140.

Tan, M., Pang, R., Le, Q.V., 2020. EfficientDet: Scalable and efficient object detection. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 10778–10787. http://dx.doi.org/10.1109/CVPR42600.2020.01079.

Thomas, P.A., Barr, J., Balaji, B., White, K., 2017. An open source framework for tracking and state estimation ('Stone Soup'). In: Kadar, I. (Ed.), Signal Processing, Sensor/Information Fusion, and Target Recognition XXVI. Vol. 10200. SPIE, International Society for Optics and Photonics, pp. 62–71. http://dx.doi.org/10.1117/12.2266249.

Vasstein, K., 2021. A high fidelity digital twin framework for testing exteroceptive perception of autonomous Vessels. Norwegian University of Science and Technology, URL: https://hdl.handle.net/11250/2781031.

Williams, J.L., 2015. Marginal multi-bernoulli filters: RFS derivation of MHT, JIPDA, and association-based member. IEEE Trans. Aerosp. Electr. Syst. 51 (3), 1664–1687. http://dx.doi.org/10.1109/TAES.2015.130550.

Wilthil, E.F., Flåten, A.L., Brekke, E.F., 2017. A target tracking system for ASV collision avoidance based on the PDAF. In: Fossen, P., Nijmeijer (Eds.), Sensing and Control for Autonomous Vehicles. Vol. 474. Springer, Alesund, Norway, pp. 269–288. http://dx.doi.org/10.1007/978-3-319-55372-6_13.

Wolf, M.T., Assad, C., Kuwata, Y., Howard, A., Aghazarian, H., Zhu, D., Lu, T., Trebi-Ollennu, A., Huntsberger, T., 2010. 360-Degree visual detection and target tracking on an autonomous surface vehicle. J. Field Robotics 27 (6), 819–833. http://dx.doi.org/10.1002/rob.20371, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/rob.20371. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.20371.

Zhang, Z., 2000. A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. 22 (11), 1330–1334. http://dx.doi.org/10.1109/34.888718.

Øystein K. Helgesen is a Ph.D. candidate at the Department of Engineering Cybernetics at the Norwegian University of Science and Technology (NTNU). His main research focus is sensor fusion and situational awareness for autonomous ferries as part of the Autoferry project.

Kjetil Vasstein is a Ph.D. candidate at the Department of Engineering Cybernetics at the Norwegian University of Science and Technology. Kjetil's main focus is within the fields of Cybernetics, sensor fusion, and the use of video games for autonomous vehicles as a part of the TRUSST project.

Edmund F. Brekke is an Associate Professor at the Department of Engineering Cybernetics, NTNU. His research is focused on estimation with applications in target tracking, navigation, and simultaneous localization and mapping (SLAM). He is a project manager in the Autosit project and its predecessor, Autosea, as well as a key researcher in the Autoferry project and a work package leader in SFI Autoship.

Annette Stahl is an Associate Professor at the Department of Engineering Cybernetics, NTNU. Her research is focused on computer vision applications and environmental perception utilized by various sensor platforms enabling autonomy. She is an affiliated scientist of the Centre of Excellence - NTNU AMOS, a researcher at the Centre for Research-based Innovation (SFI) AutoShip, and Head of the Robotic Vision Group at NTNU.