



# Experimental validation of camera-based maritime collision avoidance for autonomous urban passenger ferries

Ø.K. Helgesen<sup>1</sup> E.H. Thyri<sup>2</sup> E.F. Brekke<sup>1</sup> A. Stahl<sup>1</sup> M. Breivik<sup>1</sup>

<sup>1</sup>Department of Engineering Cybernetics, Norwegian University of Science and Technology (NTNU), NO-7491 Trondheim, Norway. E-mail: {oystein.k.helgesen, edmund.brekke, annette.stahl, morten.breivik}@ntnu.no

<sup>2</sup>Zeabuz AS, NO-7042 Trondheim, Norway. E-mail: emil.thyri@zeabuz.com

---

## Abstract

Maritime collision avoidance systems rely on accurate state estimates of other objects in the environment from a tracking system. Traditionally, this understanding is generated using one or more active sensors such as radars and lidars. Imaging sensors such as daylight cameras have recently become a popular addition to these sensor suites due to their low cost and high resolution. However, most tracking systems still rely exclusively on active sensors or a fusion of active and passive sensors. In this work, we present a complete collision avoidance system relying solely on camera tracking. The viability of this autonomous navigation system is verified through a real-world, closed-loop collision avoidance experiment with a single target in Trondheim, Norway in December 2022. Accurate tracking was established in all scenarios and the collision avoidance system took appropriate actions to avoid collisions.

*Keywords:* Maritime autonomy, target tracking, collision avoidance, daylight cameras, full-scale experiments, autonomous surface vehicle, autonomous urban passenger ferries

---

## 1 Introduction

Marine applications of autonomous systems are increasing rapidly. Urban passenger ferries (Brekke et al., 2022), local package delivery<sup>1</sup> and cargo transport<sup>2</sup> are all examples of recent applications of autonomous surface vessels that are close to realization. For safe, autonomous operation to be possible, these vessels must rely on a multitude of sub-components composing a complete autonomy system where collision avoidance (COLAV) and situational awareness (SITAW) are among the principal components.

The task of collision avoidance comes down to maneuvering so as to not collide with static obstacles

or other vessels, denoted target ships. Simultaneously, the autonomous vessel, denoted ownship, should proceed along its transit route. Therefore, collision avoidance systems usually include some degree of guidance or trajectory planning. Ensuring collision avoidance in a dynamic maritime environment requires consideration of a large variety of factors like dynamic and static obstacles, traffic rules, convergence towards the destination, and energy efficiency, and it is therefore not easily solved by a single, monolithic algorithm.

Therefore, it is common to delegate the planning obligations on three or more collision avoidance algorithms in what is often referred to as a hybrid collision avoidance system, where a deliberate planning method considers long-term or global objectives, while a reactive algorithm handles short-term objectives such as collision avoidance and path or trajectory following (Eriksen, 2019).

---

<sup>1</sup><https://roboat.org>

<sup>2</sup><https://www.yara.com/corporate-releases/yara-to-start-operating-the-worlds-first-fully-emission-free-container-ship/>

An example of such a hybrid system is presented in (Bitar et al., 2019), where an optimization-based global path planner that only considers static obstacles is paired with a short-term model-predictive control-based planner for making local adjustments to the global trajectory in order to avoid collision with other vessels. This approach allows for reaping the benefits of a computationally costly long-term planner while maintaining a responsive behavior to local conditions.

Examples of short-term or reactive collision avoidance algorithms include the velocity obstacle method in (Kuwata et al., 2014). This is a partially rules-compliant method that applies a velocity obstacle to restrict the reference velocity vector to a set of collision-free velocities. The branching course model predictive control by Eriksen et al. (2019) simulates a finite set of maneuvers over a short-term time horizon and evaluates each resulting trajectory candidate by a cost function. Another example of short-term collision avoidance in confined waters is presented in (Thyri et al., 2020) where the collision avoidance problem is reduced to a velocity planning problem along a pre-defined path.

Traditionally, situational awareness in the maritime domain has primarily been based on single-sensor systems such as radar. Examples of this include (Schuster et al., 2014) which implemented a radar-based situational awareness system with a low-cost maritime radar, and (Fowdur et al., 2021) which presents a radar-based system for extended object tracking. The authors of (Wilthil et al., 2017) describe a complete radar-based tracking pipeline from raw sensor data to the resulting state estimates of the tracking process.

In recent years, due to increased applications in confined waters and influence from the automotive industry, maritime situational awareness has seen greater use of alternative sensors such as lidars and cameras and heterogeneous sensor fusion where multiple sensors are combined in a single system for greater robustness and reliability. Camera-based tracking systems include (Schöller et al., 2020), which proposes a camera-based system based on tracking features from a neural network detection process, (Wolf et al., 2010) which utilizes a 360° camera system and (Helgesen et al., 2023), which combines range estimation with multi-camera fusion. Cameras also frequently appear in sensor fusion systems such as (Cormack et al., 2020), which combines radar with infrared cameras

for multi-target tracking or Helgesen et al. (2022) which demonstrates a sensor fusion system combining radar, lidar, infrared, and daylight cameras. Another popular sensor is the automatic identification system which is combined with radar in (Gaglione et al., 2018).

In this work, we propose a novel system for autonomous navigation in confined waters based solely on cameras for situational awareness. The system combines the collision avoidance method of Thyri et al. (2020) with the camera-based tracking system of Helgesen et al. (2023), creating what is to the authors' knowledge the first purely camera-based maritime collision avoidance system described in scientific literature. Using the *milliAmpere 2* urban autonomous passenger ferry, the system is verified in a closed-loop experiment with a single target. The autonomy system itself is described in sections 3 and 4 for the SITAW and COLAV parts respectively, while section 5 details the experimental verification. System performance during this verification is analyzed in section 6 with concluding remarks in sections 7 and 8. A video of the experimental validation is available.<sup>3</sup>

## 2 Autonomy test platform

The *milliAmpere 2* (MA2), see Figure 1 and Table 1, is a production-ready autonomous urban ferry prototype developed by the Autoferry project<sup>4</sup> at NTNU. Designed to carry 12 passengers, MA2 is highly maneuverable due to its quad thruster configuration. Sensor-wise, MA2 is equipped with a single maritime radar, two lidars on the diagonal of the ferry, and eight RGB cameras mounted to the sides of the front and rear hatch. MA2 conducted a successful multi-week trial operation<sup>5</sup> in the autumn 2022 in Trondheim, Norway, utilizing lidars and radar for situational awareness.

<b>Length</b>	8.65m
<b>Beam</b>	3.5m
<b>Displacement</b>	6 tonnes
<b>Propulsion</b>	4 x 10kW azimuth thrusters
<b>Operational speed</b>	3 knots
<b>Batteries</b>	Lead-acid 48kWh

Table 1: *MilliAmpere 2* specifications.

<sup>3</sup>Video output from the experimental validation is available at <https://youtu.be/rXJHpo0UMck>

<sup>4</sup><https://www.ntnu.edu/autoferry>

<sup>5</sup><https://norwegianscitechnews.com/2022/09/ntnu-trials-worlds-first-urban-autonomous-passenger-ferry/>



Figure 1: *MilliAmpere 2*, the experimental platform used in this work, crossing the Canal in Trondheim. Photo by Mikael Sætereid / NTNU.

### 3 Situational awareness

The situational awareness system is responsible for tracking dynamic objects in the vessel’s vicinity using exteroceptive sensors, a process known as target tracking, which the collision avoidance system can then use for navigation. Most autonomous platforms are equipped with a wide range of sensors which are combined in a sensor fusion system, however, in this work we focus only on imaging sensors in the form of RGB cameras. This section details the individual components that make up the situational awareness system, illustrated in Figure 2.

#### 3.1 Detection system

Due to their passive nature and high resolution, cameras require more advanced processing with additional stages compared to active sensors such as lidars to generate measurements for the tracking system. In this section, we present a brief overview of the detection pipeline, detailed in Helgesen et al. (2023), as implemented on MA2.

##### 3.1.1 Image processing

Images from the cameras are supplied as raw Bayer images with only a single channel. Compared to sending full three-channel RGB images this saves bandwidth with no information loss but does require an additional processing step to demosaic the images, recovering the color information encoded in the single channel. Once completed, images are then corrected for lens distortion

using pre-defined calibration parameters according to Zhang (2000).

##### 3.1.2 Detection

A Yolo v4 (Bochkovskiy et al., 2020) deep-learning detector operating on the color and distortion-corrected images is responsible for the actual detection stage in the pipeline, converting image data into bounding boxes, yielding the position of a target in pixel coordinates as seen in Figure 3.

##### 3.1.3 Range estimation

These bounding boxes are then georeferenced, utilizing implicit information about target elevations. The standard pinhole camera model (Zhang, 2000) describes the projection of 3D world points from the camera frame  $c$  which is centered in the camera aperture with axes aligned with the image plane to individual pixels in the resulting image. This model depends only on the intrinsic parameters of the camera given by the intrinsic matrix  $\mathbf{K}$ :

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

where  $f_x$  and  $f_y$  are the focal lengths of the lens along the  $x$  and  $y$  axis of the image plane and  $c_x$  and  $c_y$  the  $x$  and  $y$  coordinates of the principal point, the intersection between the principal axis and the image plane. All variables have the unit  $\frac{\text{pixel}}{\text{m}}$ .

In most applications, a world frame is used to describe the position of objects, requiring an additional transformation consisting of a rotation and translation before points can be projected into the image. This transformation is given by the extrinsic parameters consisting of the rotation matrix  $\mathbf{R}_w^c$  and translation vector  $\mathbf{t}_w^c$ , which describe the camera rotation and pose relative to the world frame  $w$ .

In the system presented in this work, a local north-east-down (NED) frame (Fossen, 2021) is used for target tracking. This requires a dynamic transform from the NED origin to the vessel center, supplied by the navigation system of the ownship, and a static transform from the vessel center to the camera frames. Using heterogeneous coordinates the transformation can be combined into the camera matrix  $\mathbf{P}$  given by

$$\mathbf{P} = \mathbf{K}[\mathbf{R}_w^c | \mathbf{t}_w^c] = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \quad (2)$$

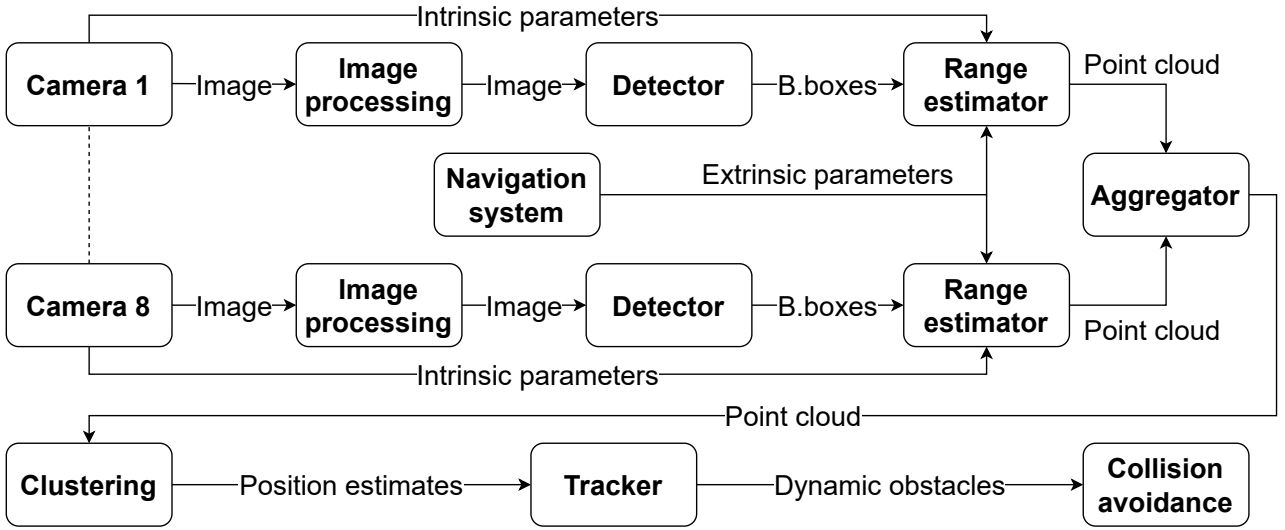


Figure 2: Situational awareness system block diagram.

where  $p_{ij}$  are individual elements of the matrix. The transformation from 3D world points,  $\mathbf{x}^w = [x^w \ y^w \ z^w]^\top$ , to 2D image points,  $\mathbf{x}^p = [x^p \ y^p]^\top$ , is then

$$s \begin{bmatrix} \mathbf{x}^p \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} \mathbf{x}^w \\ 1 \end{bmatrix} \quad (3)$$

where  $s$  is a scale factor given by the depth of the point in the camera frame. Assuming the target elevation is fixed to the ocean surface allows this model to be reversed, resulting in the position estimate

$$\begin{bmatrix} x^w \\ y^w \end{bmatrix} = \begin{bmatrix} x^p p_{31} - p_{11} & x^p p_{32} - p_{12} \\ y^p p_{31} - p_{21} & y^p p_{32} - p_{22} \end{bmatrix}^{-1} \begin{bmatrix} z^w (p_{13} - x^p p_{33}) + p_{14} - x^p p_{34} \\ z^w (p_{23} - y^p p_{33}) + p_{24} - y^p p_{34} \end{bmatrix}. \quad (4)$$

For each bounding box, this is repeated for both the left and right corners, yielding an estimate of the target extent. Points are then generated between these extremes using linear interpolation. These points are then aggregated into a single point cloud in the world frame containing points from all eight cameras.



Figure 3: Detection output from the Yolo v4 on an image from the experimental validation in the Canal in Trondheim. The algorithm detects four leisure boats docked along the Canal.

### 3.1.4 Detection filtering

The operating environment of MA2 is littered with docked boats along the edges of the Canal. If these detections were included in the tracking process, the computational cost would explode, possibly making real-time operation impossible. To correct this, we implement a filtering step that utilizes a pre-generated occupancy grid with cell values based on whether we want to track objects present in them or not. The grid itself is based on land maps supplied by the Norwegian

mapping authority with some additional masking that covers the floating docks along the Canal. If a point from the range estimation is determined to fall into one of these cells, the point is removed, leaving us with only detections that could originate from targets in the Canal.

### 3.1.5 Clustering

The final step in the pipeline fuses the filtered points into a single measurement per target, as assumed by the tracking system, using clustering. This approach removes the bias induced by targets being partially present in multiple images, resulting in two detections along the front and aft of the target, by merging the two partial detections as seen in Figure 4. This results in a single position estimate centered around the middle of the target which is then sent into the tracking system.

## 3.2 Tracking system

The tracking system itself is based on the IDPA (Muscicki et al., 1992) single target tracker which was successfully used during a three-week public trial operation of MA2<sup>6</sup> and has therefore been stress-tested over long periods of time in real-world conditions.

### 3.2.1 Motion model

An accurate motion model is a key component for reliable autonomous operations. This model is used to predict future target states, forming the basis of collision avoidance decision-making. Due to the limited maneuvering room in the operational area of MA2, a constant velocity model is used. This model assumes targets have constant velocities with acceleration modeled as Gaussian white noise. Target states are given by

$$\mathbf{x}^w = \begin{bmatrix} x^w \\ y^w \\ v_x^w \\ v_y^w \end{bmatrix} \quad (5)$$

where  $x^w, y^w$  are positions and  $v_x^w, v_y^w$  velocities in the world frame. For non-continuous applications, this model is discretized as

$$\mathbf{x}_k^w = \mathbf{F}\mathbf{x}_{k-1}^w + \mathbf{v}_k \quad \mathbf{v}_k \sim \mathcal{N}(0, \mathbf{Q}) \quad (6)$$

where  $\mathbf{x}_k$  is the state at time-step  $k$ ,  $\mathbf{v}_k$  the discretized process noise with covariance  $\mathbf{Q}$ , and  $\mathbf{F}$  the state transition matrix.

<sup>6</sup><https://norwegianscitechnews.com/2022/09/ntnu-trials-worlds-first-urban-autonomous-passenger-ferry/>

### 3.2.2 Sensor model

After the clustering-based camera fusion is applied the camera detection pipeline outputs Cartesian detections. This yields the same measurement function as used for the active sensors in (Helgesen et al., 2022)

$$f_z^c(\mathbf{x}_k^w) = \begin{bmatrix} x_w \\ y_w \end{bmatrix} + \mathbf{w}_k \quad \mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}) \quad (7)$$

which results in a 2D position measurement where  $\mathbf{w}_k$  is zero-mean Gaussian white noise with covariance  $\mathbf{R}$ .

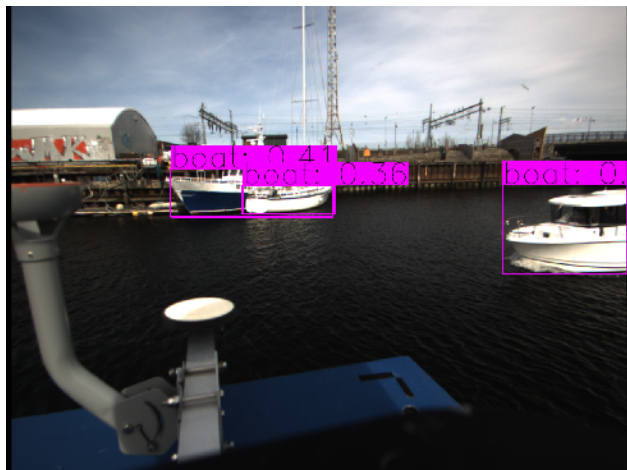
### 3.2.3 Track management

Track management is the component of the tracking system responsible for establishing new tracks and terminating existing ones. Accurate track management should establish new tracks with a minimal delay once detections are available and maintain the existence of valid tracks while minimizing the number of false tracks. With existence-based trackers such as the IPDA, tracks are often split into two categories, preliminary and confirmed, where only the confirmed tracks are considered by the COLAV system.

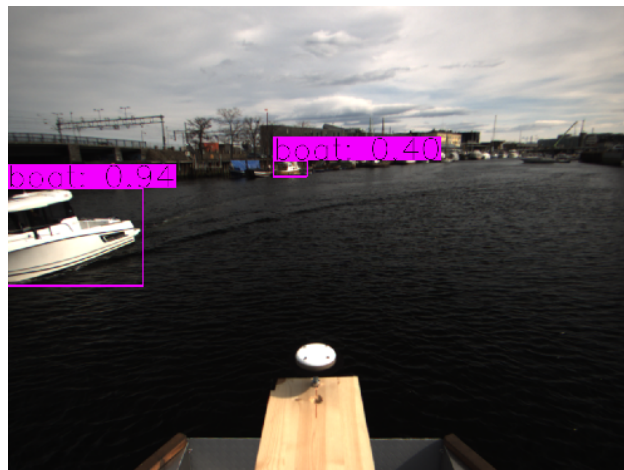
New tracks are established on any measurement not associated with a preliminary or confirmed track. These tracks are established at a pre-determined existence probability of 0.2 as preliminary tracks. Promotion to confirmed track status happens if the existence probability grows above a threshold of 0.9. Existing tracks are removed if the existence probability falls below a threshold of 0.1 or if the position covariance exceeds a standard deviation of 50m. Both of these indicate that the track is unlikely to exist, either directly or indirectly. Terminating these tracks reduces computational complexity and contributes toward more consistent system behavior. The parameter values for these thresholds are identical to the MA2 trial operation system.

## 4 Collision avoidance

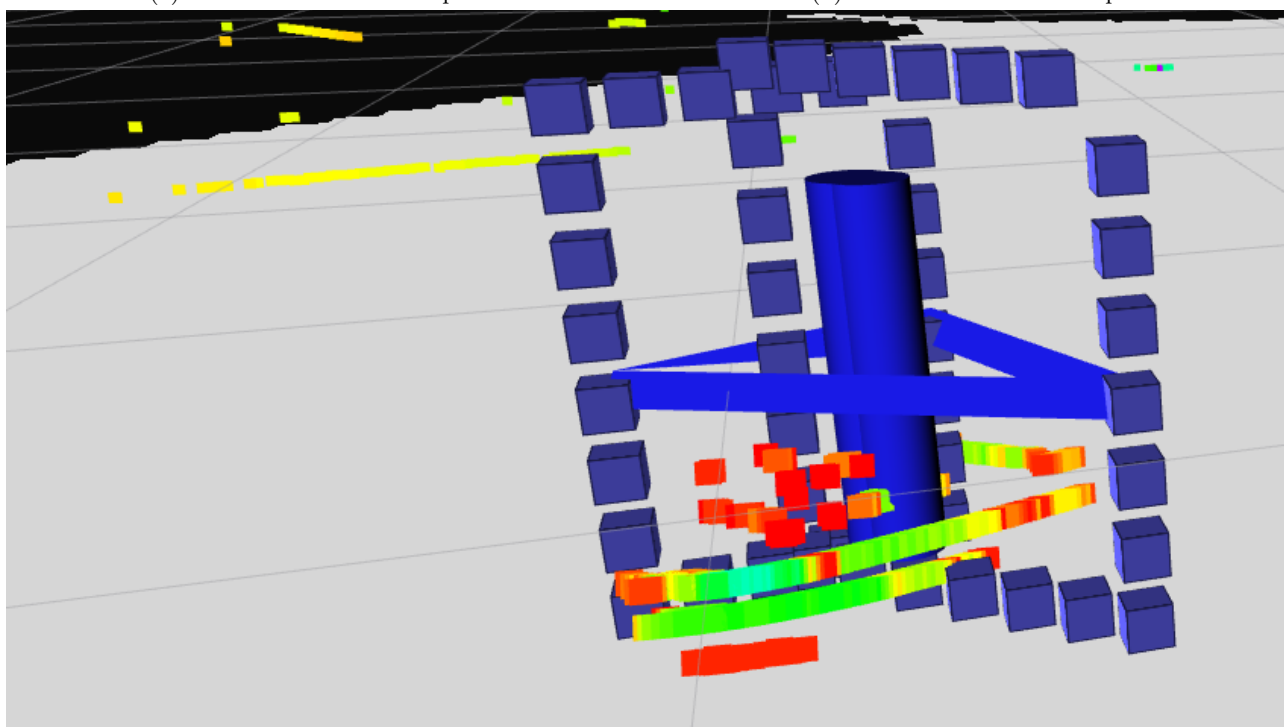
The collision avoidance system applied in this work is based on the trajectory planner proposed in (Thyri et al., 2020). The method applies the principles of path-velocity decomposition, where first a path that is collision-free with static obstacles is determined. Then, a velocity profile is planned for that path so that the resulting trajectory is collision-free with dynamic obstacles. In this work, a nominal path is predefined and applied in all scenarios. The path is collision-free with static obstacles, and goes across the Canal, as illustrated by the red path of the ownship in Figure 10a.



(a) Camera 1 detection output.



(b) Camera 2 detection output.



(c) Position estimates of the partial bounding boxes from Figures 4a and 4b (blue squares) and the resulting cluster outline (deep blue polygon with cylindrical center marking), overlaid on lidar data (colored dots).

Figure 4: Fusion of detections from a target partially visible in two camera frames on the *milliAmpere 1* ferry from a previous data collection.

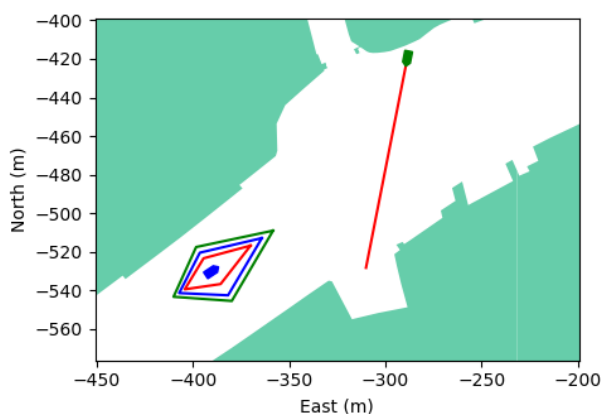


Figure 5: A blue target ship approaching a nominal path across the Canal for the green ownship. The vessel safety domain is shown as three polygons in red, blue, and green, representing the ROC, HPR, and LPR respectively.

Collision avoidance with target ships maneuvering in the Canal, therefore, becomes a velocity planning problem for the nominal path, which is solved by the following steps:

1. A safety domain is assigned to the target ship vessel.
2. The safety domain is transformed onto the path-time space.
3. A graph traversing the path-time space is built and searched to find a collision-free velocity profile for the path.

The steps are given in more detail in the following sections.

#### 4.1 Safety domain

The safety domain is composed of three polygons, as shown in Figure 5. The red polygon represents the region of collision (ROC) which is formulated so that if the position of the ownship does not violate the ROC, the vessels are collision-free. Additionally, the domain comprises the high- and low-penalty regions denoted HPR and LPR respectively. The HPR and LPR can be traversed by the ownship, however, a cost is assigned to the segments of the graph that is traversing the HPR and LPR.

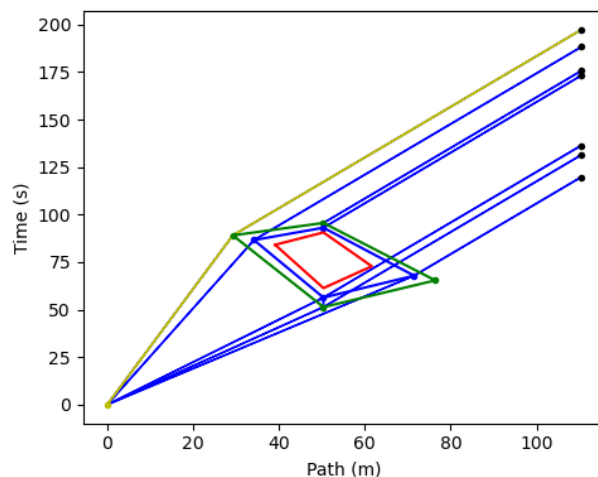


Figure 6: Path-time space for the scenario in Figure 5. The transformed safety domain has vertices at the corners of the HPR and LPR. The root node is at the origin, and the end vertices are along the right side of the path-time space. The minimum-cost velocity profile is shown in yellow.

#### 4.2 Transformation to path-time space

Once the safety domain is assigned, the domain is transformed onto a space spanned by the nominal path and time. In the path-time space, the transformed domain represents the safety domain’s occupation of the path in time. In this work, the nominal path that is applied is defined by a set of waypoints connected by straight line segments. By assuming that the target ship will maintain a constant velocity and heading until its safety domain has cleared the path, the transformation to path-time space can be solved by a set of linear equations (Thyri et al., 2020). In Figure 6 the transformed safety domain of the target ship vessel onto the red path in Figure 5 is shown. The transformed ROC, HPR, and LPR are shown as red, blue, and green polygons respectively.

#### 4.3 Building and searching the graph

Once the safety domain is transformed, a directed graph is constructed in the path-time space. The graph is constructed from a set of vertices. Each vertex that is added to the graph is attempted connected with the existing vertices by a feasible edge. An edge connecting two vertices corresponds to traversing a segment of the path at a fixed speed given by the slope of the edge. An edge is considered admissible if the following three criteria are met:

1. The edge speed is within the feasible range of the ownship.
2. The edge is moving forward in time.
3. The edge is not intersecting an ROC.

The set of vertices that are added to the graph are

- A root vertex at the current path-time coordinate of the ownship.
- Delay vertices with the same path coordinate as the root vertex but shifted along the positive time-axis. Edges connecting the root-vertex to delay vertices have zero speed and correspond to holding position.
- Obstacle vertices are added at the corners of the transformed HPR and LPR.

Finally, the graph is connected to the end of the path by attempting to add an edge between each vertex already in the graph and a vertex with a path coordinate corresponding to the end of the path and an edge speed equal to the desired transit speed for the crossing. A graph for the scenario in Figure 5, where the ownship is at the start of the path is shown in Figure 6.

A cost is calculated for each edge in the graph, where deviation from the desired transit speed and intersection with the HPR and LPR are penalized. This facilitates low-cost velocity profiles that are biased towards the desired transit speed and against domain violation. The minimum cost collision-free velocity profile is found by searching the graph by Dijkstra's algorithm.

#### 4.4 Replanning

When transforming the safety domain onto the path-time space, it is assumed that the target maintains a constant velocity. This constant velocity assumption makes the precision of the transformation suffer when violated, e.g. target maneuvering, and from uncertainty in the tracker estimates. Therefore, the safety domain is periodically re-transformed to the path-time space based on the most recent tracking data. If a collision conflict between the new ROC and the current velocity profile is identified, a new collision-free velocity profile is planned.

## 5 Experimental validation

Proper validation of the situational awareness system requires a platform capable of autonomous operation in addition to specifically designed scenarios with a controllable target ship that would result in collisions if no



Figure 7: Buster XL target vessel in the Canal in Trondheim. Photo by Karl Edvard Dalhaug.

action is taken by the platform. This section presents the target vessel and the scenarios used to validate system performance.

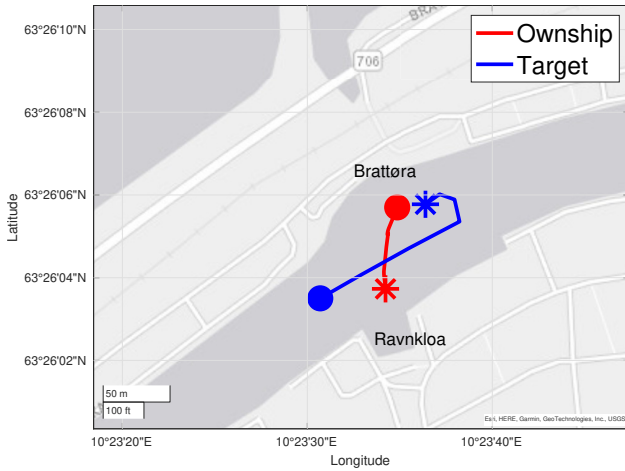
### 5.1 Target vessel

The target vessel used in this work is a Buster XL aluminum leisure boat, Figure 7, equipped with a Garmin eTrex 10 GNSS receiver for position logging. With a length of 6.05m and a width of 2.2m, this vessel is on the lower end of the average size that we expect to encounter in the Canal, especially when vessel height is also considered. Compared to other vessels used previously, such as *Havfruen* in Helgesen et al. (2022), both the size and contrast between the vessel hull and the water are lower. This could negatively influence detection performance, especially for a camera-based system.

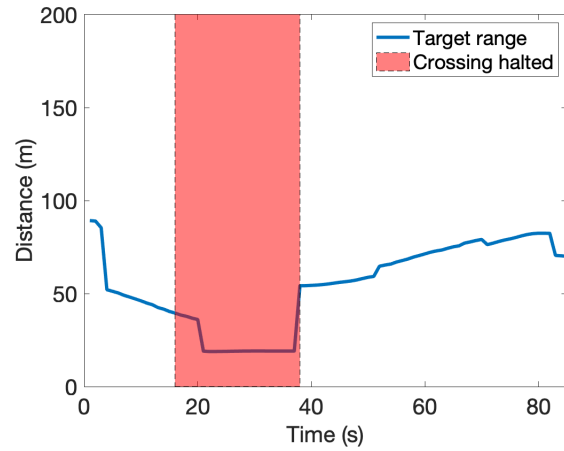
### 5.2 Scenario description

A total of six crossings of the Canal in Trondheim, identical to the trial operation route, were conducted, where each crossing contained at least one interference from the target vessel that would have resulted in a collision if no action were to be taken by the autonomy system. Both weather and lighting conditions proved challenging, the experiments were performed in December 2022 in Trondheim where daylight hours are limited and light intensity low. An additional challenge was provided by the snowy weather conditions which brought with them significant cloud coverage and the potential for camera obscuration due to snow hitting the lens.

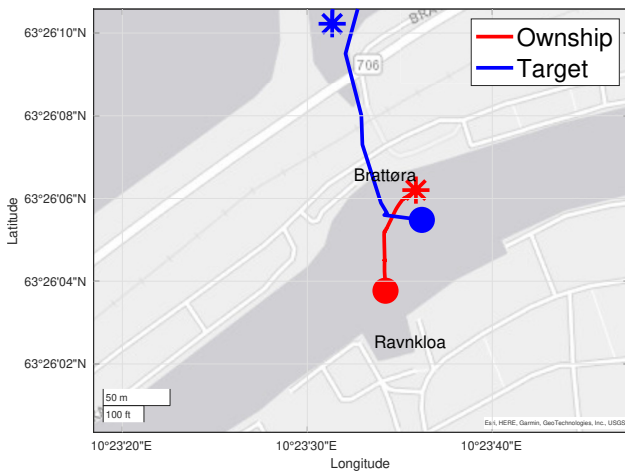




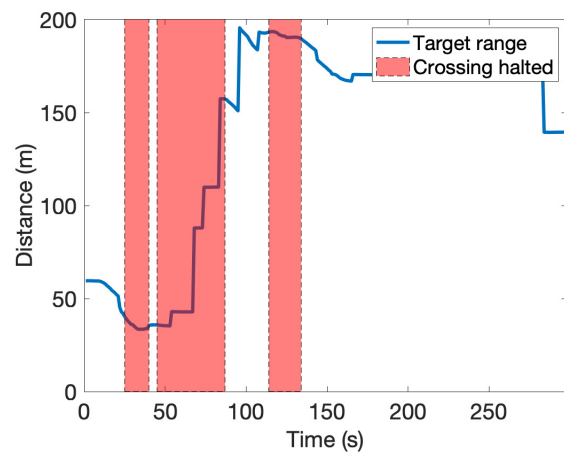
(a) Crossing 1, ownship and target maneuvers.



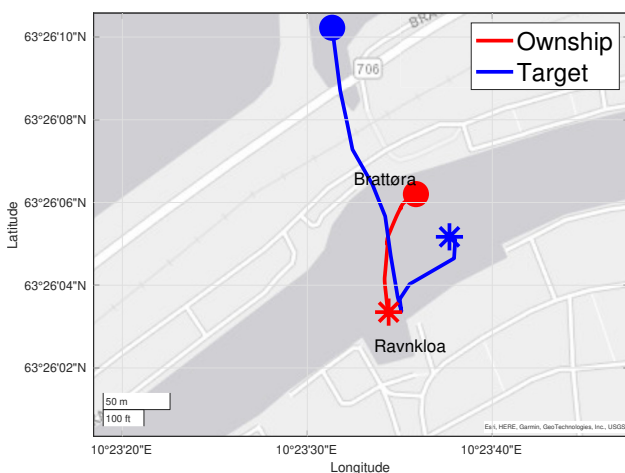
(b) Crossing 1, target range and COLAV behaviour.



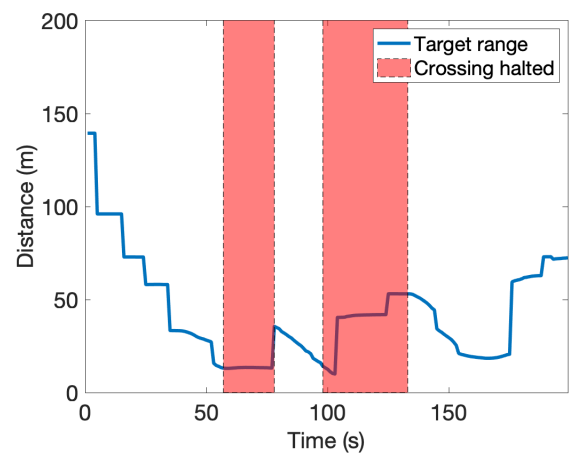
(c) Crossing 2, ownship and target maneuvers.



(d) Crossing 2, target range and COLAV behaviour. The final COLAV maneuver was induced by a temporary false track.



(e) Crossing 3, ownship and target maneuvers.



(f) Crossing 3, target range and COLAV behaviour.

Figure 8: Crossings 1-3. Circles signify starting locations, stars signify end locations.

## 6 Experimental results

In this section, we present the results of the experimental validation described in section 5, including crossings and target maneuvers, and the resulting autonomy system behaviour. All maps shown in this section, e.g. Figure 8a, have axes aligned with the global NED frame, i.e. north up and east to the left. System behaviour plots, e.g. Figure 8b, illustrate when the COLAV system halted transit due to collision risk.

### 6.1 Crossing 1

In crossing 1, Figure 8a, MA2 starts at the north end of its route on the Brattøra side of the Canal. The target travels from west towards east, crossing in front of the ferry during the mid-point of its route. A single maneuver was performed by the collision avoidance system when target range approached 20m, Figure 8b. The ownship was then stationary for 15s as the target vessel passed to the front.

### 6.2 Crossing 2

Crossing 2, Figure 8c, begins on the south side of the Canal with MA2 situated a few meters out from its docking location. The target starts out stationary in the middle of the intended path, resulting in a collision if no action is taken. The target then continues underneath the north-side bridge. This resulted in a single maneuver at a range of 30m where the ownship stopped completely. A momentary track loss was experienced by the SITAW system, however, the ownship remained stationary during this period and a track was quickly re-established. Once the vessel continued underneath the bridge the crossing was resumed until a false track led to a very brief stop before docking, see Figure 8d

### 6.3 Crossing 3

This crossing adds additional difficulty by having the target appear from underneath a wide bridge with poor lighting conditions, see Figure 8e. The ownship is situated in the docking adapter just to the side of the bridge with the target appearing at a range of less than 50m. Once visible to our eyes, a crossing is initiated that would result in a collision during the initial phase. The target continues towards the dock on the other side, leading the ownship by some meters. Two significant COLAV maneuvers were performed due to this, Figure 8f. The first one occurred at a range of 20m when the target crossed the trajectory of the ownship, resulting in a brief stop. The camera image in Figure 9 visualizes this situation. As the target continued



Figure 9: Crossing 3, target intersecting ownship trajectory, inducing COLAV maneuver.

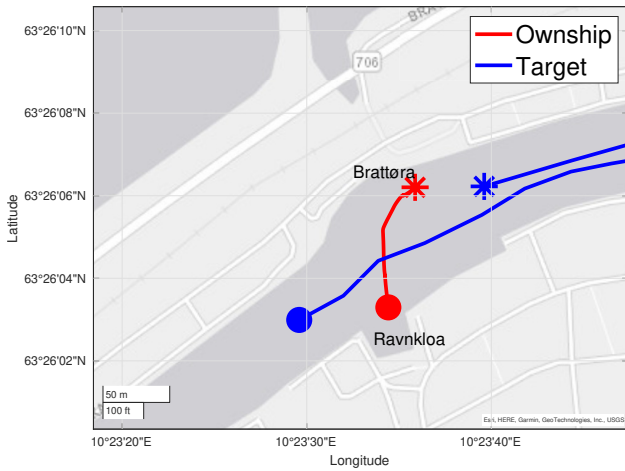
parallel to the ownship trajectory a second maneuver was performed when the tracking system reported that safety margins would be violated if the current trajectory was maintained.

### 6.4 Crossing 4

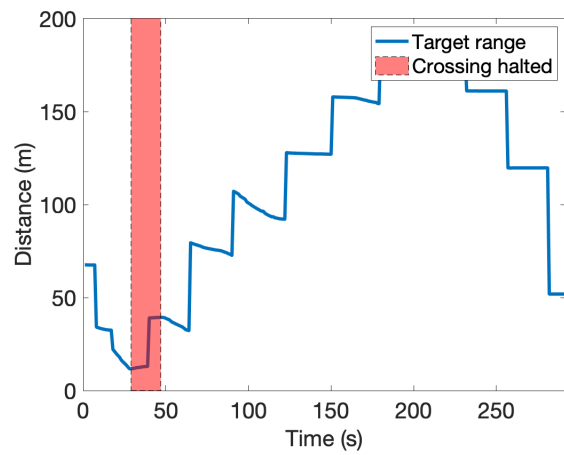
Similar to crossing 2, this scenario has the ownship traveling south to north. A single intersection is present, however, this time the target is traveling west to east and the intersection happens at a much earlier phase. A visualization is shown in Figure 10a. For this crossing, only a single maneuver was required when the target passed directly in front of the moving ownship as seen in Figure 10b. Crossing was then resumed swiftly with no further interruptions.

### 6.5 Crossing 5

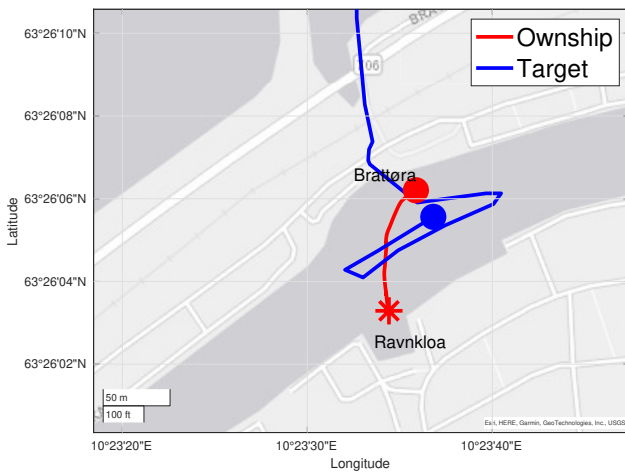
Crossing 5, Figure 10c, is a difficult scenario for the tracking system due to target vessel maneuvers. After an initial intersection roughly in the middle of the crossing, the target performs a u-turn to intersect the ferry twice. This challenges the constant velocity model used in the tracker and could potentially result in track loss. This crossing was designed to require multiple maneuvers from the COLAV. The first intersection of the target and the ownship's intended path resulted in the COLAV system temporarily halting transit. Once the target was clear, the crossing was resumed automatically. A second maneuver was performed in the middle of the crossing when the target had turned around to cross the ownship trajectory a second time. Again, the target was successfully tracked



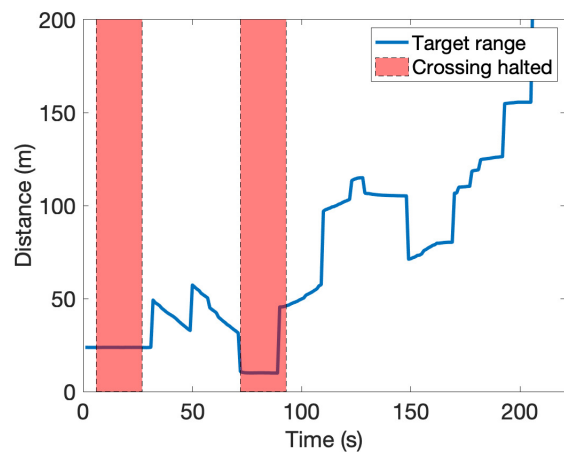
(a) Crossing 4, ownship and target maneuvers.



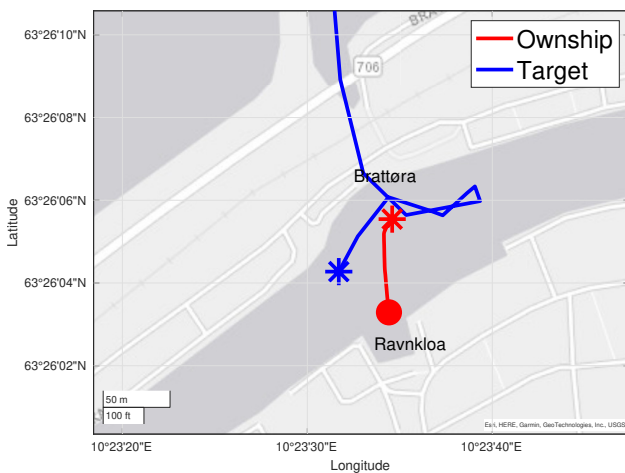
(b) Crossing 4, target range and COLAV behaviour.



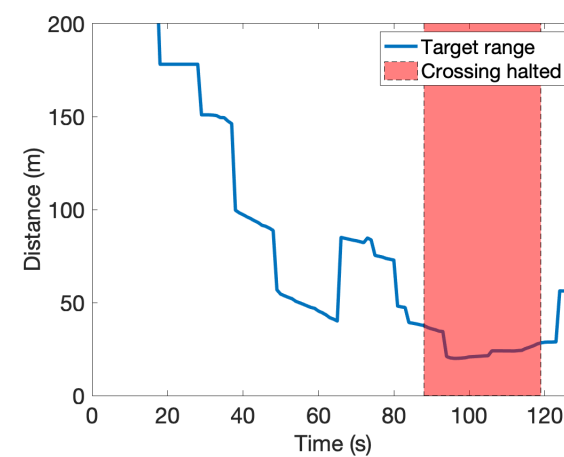
(c) Crossing 5, ownship and target maneuvers.



(d) Crossing 5, target range and COLAV behaviour.



(e) Crossing 6, ownship and target maneuvers.



(f) Crossing 6, target range and COLAV behaviour.

Figure 10: Crossings 4-6. Circles signify starting locations, stars signify end locations.

and the autonomy system was able to avoid colliding, shown in Figure 10d.

## 6.6 Crossing 6

This crossing, Figure 10e, repeats the target appearing from underneath the bridge. However, this time the ownship starts out at the opposite end of the Canal allowing a much greater safety margin when the target appears. A crossing is initiated when the target can be seen underneath the bridge, the target then turns east continuing along the Canal well clear of the ownship. After a short period of time, the target again performs a u-turn, this time intersecting the ferry at a very close range along its direction of travel just before the docking sequence is initiated on the north side. This crossing was designed to require multiple maneuvers from the COLAV. The first intersection of the target and the ownship’s intended path resulted in the COLAV system temporarily halting transit. Once the target was clear, the crossing was resumed automatically. A second maneuver was performed in the middle of the crossing when the target had turned around to cross the ownship trajectory a second time. Again, the target was successfully tracked and the autonomy system was able to avoid colliding, shown in Figure 10d.

## 7 Discussion and future work

While the experimental validation of this autonomy system was successful, some issues were uncovered, see Table 2 for an overview. With eight Ethernet cameras operating at 5 Hz, MA2 consumes large amounts of network bandwidth. During the experimental validation, this became an issue causing approximately 40% of the images to be dropped. This did cause premature track death in some of the preliminary tracking phases due to a series of missing detections right after track establishment. Restoring full functionality would allow more rapid track establishment, enabling faster reaction from the COLAV system, and increasing the safety margins of the system for targets appearing close to the vessel such as in Crossing 3.

Here, the autonomy system initiated a crossing even though the target was set to intersect the ownship trajectory almost immediately after departure. The system did eventually stop with some margin, however, the crossing would in all likelihood have been delayed until the target was clear of the planned path if a confirmed track had been established. The wide-angle lenses equipped on the ferry also reduce the maximum detection range of the camera detector,

Work item	Impact
Restore full camera functionality	High
Increase camera elevation	High
Improved target trajectory predictions	High
Integration with lidar and radar in a sensor fusion system	High
Long range detections	Medium
Alternative range estimation methods	Medium
Traffic regulations compliance	Medium

Table 2: Improvements and future work. Impact is only considered for the current MA2 operational domain.

trading this for increased coverage around the ferry. In the current operating environment, this does not pose an issue due to the low speed of both targets and ownship. For better generalization, the system should be able to track targets at a greater distance, especially for applications where ownship or target speeds are higher, requiring larger maneuvering margins.

In terms of estimation accuracy, the main limiter is the low altitude of the current camera mounting locations. This increases the range estimates sensitivity to noise, both in detection and navigation. Higher mounting locations would yield more accurate estimates, reducing the noise sensitivity of the system. Deep learning-based methods for monocular depth estimation have in recent years started appearing (Kuznietsov et al., 2017) and could be a viable alternative to the geometry-based estimation method used in this work. Advantages include the potential for range estimation for all parts of the image, including non-target areas. Targets would also receive several discrete estimation points, yielding a more accurate shape estimate than the current bounding box-based method.

On the other hand, these methods rely on large amounts of training data and can fail when encountering situations differing from this data. In addition, the computational expense is far greater which can pose an issue for real-time performance with eight cameras. Stereo vision is also a possible alternative for depth estimation. MA2 is already equipped for this with dual cameras along the main axes of travel, however, no actual stereo vision algorithm is currently implemented on the platform and the wide field of view of the cameras could limit long-distance accuracy.

For the collision avoidance performance, the precision of the planned velocity profile can be greatly increased by a more educated prediction of the future

trajectory of the target vessels. In the current prediction, a constant velocity model is applied, however, for operation in such confined areas, the presence of static obstacles makes this a weak assumption. For these canal-like areas, a possible approach could be to assume the target ship maintains a constant cross-track error to the center line of the waterway or to a path constructed from historical traffic data.

Furthermore, the collision avoidance method does not consider any traffic regulations, such as the international regulations for preventing collision at sea, to any degree. These regulations dictate the maneuvering obligations of vessels in encounters where the risk of collision is present. In particular, the rules regarding give-way crossing encounters and stand-on crossing encounters are relevant for this type of canal operation. Here, the ownship has stand-on obligations for vessels approaching from the port side, and should hence not maneuver for these vessels but instead keep a constant velocity. This can be achieved by first classifying each encounter to determine the encounter type, and then simply omitting the safety domain from the path-time space for target vessels in stand-on encounters.

## 8 Conclusion

This work has presented a novel, camera-based autonomy system for autonomous surface vehicles operating on a fixed path similar to a virtual cable ferry. The system itself consists of a situational awareness module that provides data to a collision avoidance module. Experimental validation of this system in the Canal in Trondheim, December 2022, showed that the system was able to act appropriately for a small target performing a variety of maneuvers, showcasing that cheaper imaging sensors can be a viable sensor for autonomous navigation, providing both target classification and range extension when combined with shorter range lidars in a sensor fusion system, or as a primary navigation sensor when active sensors are considered too expensive.

Several avenues for future work were identified during experimental validation, including both situational awareness accuracy and collision avoidance traffic regulations compliance.

## Acknowledgments

This work was supported in part by the Research Council of Norway through Projects 223254, 295033 and 309230, and by NTNU through the Digital transformation project Autoferry.

## References

- Bitar, G., Eriksen, B.-O. H., Lekkas, A. M., and Breivik, M. Energy-Optimized Hybrid Collision Avoidance for ASVs. In *18th European Control Conference (ECC)*. pages 2522–2529, 2019. doi:[10.23919/ECC.2019.8795645](https://doi.org/10.23919/ECC.2019.8795645).
- Bochkovskiy, A., Wang, C., and Liao, H. M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *CoRR*, 2020. abs/2004.10934. URL <https://arxiv.org/abs/2004.10934>.
- Brekke, E. F., Eide, E., Eriksen, B.-O. H., Wilthil, E. F., Breivik, M., Skjellaug, E., Øystein K Helgesen, Lekkas, A. M., Martinsen, A. B., Thyri, E. H., Torben, T., Veitch, E., Alsos, O. A., and Johansen, T. A. milliAmpere: An Autonomous Ferry Prototype. *Journal of Physics: Conference Series*, 2022. 2311(1):012029. doi:[10.1088/1742-6596/2311/1/012029](https://doi.org/10.1088/1742-6596/2311/1/012029).
- Cormack, D., Schlangen, I., Hopgood, J. R., and Clark, D. E. Joint Registration and Fusion of an Infrared Camera and Scanning Radar in a Maritime Context. *IEEE Transactions on Aerospace and Electronic Systems*, 2020. 56(2):1357–1369. doi:[10.1109/TAES.2019.2929974](https://doi.org/10.1109/TAES.2019.2929974).
- Eriksen, B.-O. H. *Collision avoidance and motion control for autonomous surface vehicles*. Ph.D. thesis, Norwegian University of Science and Technology, 2019. URL <https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2616394>.
- Eriksen, B.-O. H., Breivik, M., Wilthil, E. F., Flåten, A. L., and Brekke, E. F. The branching-course model predictive control algorithm for maritime collision avoidance. *Journal of Field Robotics*, 2019. 36(7):1222–1249. doi:<https://doi.org/10.1002/rob.21900>.
- Fossen, T. *Handbook of Marine Craft Hydrodynamics and Motion Control*. Wiley, 2021. URL <https://books.google.no/books?id=tCQqEAAAQBAJ>.
- Fowdur, J. S., Baum, M., and Heymann, F. Real-World Marine Radar Datasets for Evaluating Target Tracking Methods. *Sensors*, 2021. 21(14):4641. doi:<https://doi.org/10.3390/s21144641>.

- Gaglione, D., Braca, P., and Soldi, G. Belief Propagation Based AIS/Radar Data Fusion for Multi-Target Tracking. In *21st International Conference on Information Fusion (FUSION)*, pages 2143–2150, 2018. doi:[10.23919/ICIF.2018.8455217](https://doi.org/10.23919/ICIF.2018.8455217).
- Helgesen, Ø. K., Stahl, A., and Brekke, E. F. Maritime tracking with georeferenced multi-camera fusion. *IEEE Access*, 2023. 11:30340–30359. doi:[10.1109/ACCESS.2023.3261556](https://doi.org/10.1109/ACCESS.2023.3261556).
- Helgesen, Ø. K., Vasstein, K., Brekke, E. F., and Stahl, A. Heterogeneous multi-sensor tracking for an autonomous surface vehicle in a littoral environment. *Ocean Engineering*, 2022. 252:111168. doi:<https://doi.org/10.1016/j.oceaneng.2022.111168>.
- Kuwata, Y., Wolf, M. T., Zarzhitsky, D., and Huntsberger, T. L. Safe Maritime Autonomous Navigation with COLREGS, Using Velocity Obstacles. *IEEE Journal of Oceanic Engineering*, 2014. 39(1):110–119. doi:[10.1109/JOE.2013.2254214](https://doi.org/10.1109/JOE.2013.2254214).
- Kuznetsov, Y., Stückler, J., and Leibe, B. Semi-Supervised Deep Learning for Monocular Depth Map Prediction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2215–2223, 2017. doi:[10.1109/CVPR.2017.238](https://doi.org/10.1109/CVPR.2017.238).
- Musicki, D., Evans, R., and Stankovic, S. Integrated probabilistic data association (IPDA). In *31st IEEE Conference on Decision and Control*, pages 3796–3798 vol.4, 1992. doi:[10.1109/CDC.1992.370951](https://doi.org/10.1109/CDC.1992.370951).
- Schuster, M., Blauch, M., and Reuter, J. Collision Avoidance for Vessels using a Low-Cost Radar Sensor. *19th IFAC World Congress*, 2014. 47(3):9673–9678. doi:<https://doi.org/10.3182/20140824-6-ZA-1003.01872>.
- Schöller, F., Blanke, M., Plenge-Feidenhans', M., and Nalpantidis, L. Vision-based Object Tracking in Marine Environments using Features from Neural Network Detections. *21th IFAC World Congress*, 2020. 53(2):14517–14523. doi:<https://doi.org/10.1016/j.ifacol.2020.12.1455>. 21st IFAC World Congress.
- Thyri, E. H., Breivik, M., and Lekkas, A. M. A Path-Velocity Decomposition Approach to Collision Avoidance for Autonomous Passenger Ferries in Confined Waters. *21th IFAC World Congress*, 2020. 53(2):14628–14635. doi:<https://doi.org/10.1016/j.ifacol.2020.12.1472>.
- Wilthil, E. F., Flåten, A. L., and Brekke, E. F. A target tracking system for ASV collision avoidance based on the PDAF. In P. Fossen and Nijmeijer, editors, *Sensing and Control for Autonomous Vehicles*, volume 474, pages 269–288. Springer, Alesund, Norway, 2017. URL [http://dx.doi.org/10.1007/978-3-319-55372-6\\_13](http://dx.doi.org/10.1007/978-3-319-55372-6_13).
- Wolf, M. T., Assad, C., Kuwata, Y., Howard, A., Agazarian, H., Zhu, D., Lu, T., Trebi-Ollennu, A., and Huntsberger, T. 360-degree visual detection and target tracking on an autonomous surface vehicle. *Journal of Field Robotics*, 2010. 27(6):819–833. doi:<https://doi.org/10.1002/rob.20371>.
- Zhang, Z. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000. 22(11):1330–1334. doi:[10.1109/34.888718](https://doi.org/10.1109/34.888718).