Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 53-2 (2020) 14476–14481

# Low Altitude Georeferencing for Imaging Sensors in Maritime Tracking ⋆

**Øystein Kaarstad Helgesen* Edmund Førland Brekke* Annette Stahl* Øystein Engelhardtsen****

* *Norwegian Universtity of Science and Technology, Trondheim, Norway (e-mail: {oystein.k.helgesen, edmund.brekke, annette.stahl}@ntnu.no)*

** *DNV GL, Høvik, Norway (e-mail: oystein.engelhardtsen@dnvgl.com)*

**Abstract:** This paper presents a method for georeferencing low-altitude camera sensors, both infrared and electro-optical, in a maritime context. Accurate georeferencing require very high precision for the object pixel coordinates due to sensor resolution. To achieve this we refine the bounding boxes provided by an SSD object detector using the Sobel operator and the Hough transform. Using real world data this method is applied in a maritime tracking system based on the Joint Integrated Probabilistic Data Association method and compared to radar tracking. The georeferenced cameras surpassed radar performance in several of the benchmarks and maintained tracks with greater reliability at the cost of reduced position accuracy.

Copyright © 2020 The Authors. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0)

*Keywords:* Autonomous surface vehicles; Sensing; Marine system navigation, guidance and control

## 1. INTRODUCTION

Sensor redundancy plays an important part in safe and reliable navigation for autonomous vessels. Traditionally this has taken the form of multiple radars or in more recent years the combination of radars and lidars in a sensor fusion system. The addition of passive sensors such as cameras can further enhance the accuracy and robustness of the system. However, the lack of explicit range information presents a challenge which could significantly degrade the accuracy of the vessels tracking system.

Low cost and efficient packaging make imaging sensors attractive alternatives to more expensive, active sensors such as radars. In the maritime domain the application of these sensors for situational awareness has been the subject of considerable research focus in recent years. (Bloisi and Iocchi, 2009) demonstrated a video surveillance system for boat traffic monitoring in Venice. (Fefilatyev et al., 2010) used a buoy mounted camera to track marine vessels using a multiple hypothesis framework. Background subtraction of camera data was used to track vessel outlines in (Szpak and Tapamo, 2011).

Georeferencing, the act of associating information with geographic location, has not been discussed in comprehensive survey papers on maritime situational awareness using cameras such as (Prasad et al., 2017). Nevertheless methods using georerefencing have been examined in some previous papers. (Park et al., 2015) used a monocular camera mounted on an unmanned surface vessel to estimate target ranges based on the vertical distance between the target and the horizon. (Woo and Kim, 2016) demonstrated a vision-based collision avoidance system using georeferenced cameras on simulated data. (Helgesen et al., 2018) employed an unmanned aerial vehicle (UAV) mounted thermal camera to achieve centimeter accuracy in georeferencing static, maritime objects.

In this work we present, demonstrate and evaluate a method for extracting range information from camera data for use in a maritime target tracking system. Compared to (Park et al., 2015) our method is simpler and more robust due to eliminating the need for horizon detection. This allows applications in situations where the horizon is obscured such as urban environments or adverse weather conditions. In contrast with (Helgesen et al., 2018) we mount the cameras at low altitude more representative of vessel-mounted sensors. The position estimation method itself is similar to (Woo and Kim, 2016), however we also integrate it into a complete, state-of-the-art pipeline from detection to tracking. The tracking system is based on the Markov-chain Two version of the Joint Integrated Probabilistic Data Association (JIPDA), (Musicki and Evans, 2004), method for multi-target tracking. Real world data from infrared (IR) and electro-optical (EO) cameras are used to evaluate performance against a radar benchmark on a dataset covering both day and night conditions at ranges from 100m to 400m.

## 2. POSITION ESTIMATION

Typically it is not possible to estimate the position of an object from 2-dimensional image data without additional information, e.g. constraining the object position to a known plane. For maritime target tracking a safe assumption is that all objects of interest, excepting seaplanes, will be situated on the ocean surface which can be

⋆ This work was supported by the Research Council of Norway (NFR) through the projects 223254 and 244116/O70.

2405-8963 Copyright © 2020 The Authors. This is an open access article under the CC BY-NC-ND license.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2020.12.1449

approximated as a flat plane. By placing the camera above the ocean plane we can leverage this constraint to estimate the actual position of a target in three dimensions.

By utilizing the pinhole model and the camera calibration of (Zhang, 2000) we extract the object's bearing and elevation angle in the camera frame, creating a unit vector. The camera position and pose is used to transform this vector into a north-east-down world frame fixed at the ocean surface centered on the sensor rig. The object's position is then given by the intersection of the vector and the ocean plane.

The method can be summarized as follows:

(1) Object detection method (e.g. SSD) resulting in bounding boxes for all objects of interest.
(2) Refine the bounding box using the horizontal Sobel operator and Hough transform to find the intersection between the object and the ocean surface
(3) Use the refined bounding box position to create a vector pointing towards the object. The intersection between the vector and the ocean plane yields the object's position.

An alternative approach is to use Mask R-CNN (He et al., 2017), a deep learning detector for object segmentation, eliminating the need for bounding box refinement. This does however require specially labelled datasets with object masks and is more computationally expensive than the current method, but could result in greater accuracy.

Given a pixel position, $\mathbf{x}_P^c = [x_P^c, y_P^c]$, we first find the bearing $\theta^c$ and elevation $\varphi^c$ of the pixel in the camera frame $(c)$ relative to the image center.

$$\theta^c = \frac{x_P - R_x/2}{R_x} F_x \tag{1}$$

$$\varphi^c = \frac{y_P - R_y/2}{R_y} F_y \tag{2}$$

$R_x$ and $R_y$ denote the image resolution in pixels along the $x$ and $y$ axis while $F_x$ and $F_y$ are the corresponding fields of view (FOV) in radians of the camera. These angles are used to create a vector in the camera coordinate system pointing towards the detection, $\mathbf{v}^c$.

$$\mathbf{v}^c = [x^c, y^c, z^c] = [\tan\theta^c, \tan\varphi^c, 1] \tag{3}$$

Using the camera position and pose this vector is transformed into the world coordinate system $(w)$ by a rotation, $\mathbf{R}_c^w$, and a translation, $\mathbf{t}_c^w$, to yield $\mathbf{v}^w$,

$$\mathbf{v}^w = \mathbf{R}_c^w \mathbf{v}^c + \mathbf{t}_c^w. \tag{4}$$

The start point of the vector is in the camera center given by the translation vector. The vector end point occurs once the vector crosses the ocean plane, i.e. at $z^w = 0$. Finding the scale factor, $s$, required for the vector $\mathbf{v}^w$ to intersect is then given by the cameras elevation, $t_{c_Z}^w$, and the downwards component of the object vector, $z^w$

$$s = -\frac{t_{c_Z}^w}{z^w}. \tag{5}$$

Combining the scale factor with the object vector and the camera location yields the object's position,

$$\mathbf{x}^w = \mathbf{t}_c^w + s\mathbf{v}^w. \tag{6}$$

### 2.1 Tracking system

In the context of autonomous vessels and collission avoidance the end goal of a sensing system is accurate state estimates. The presence of clutter, sensor noise and multiple targets require specialized state estimation methods to yield optimal results. Due to this we have chosen to integrate and evaluate the georeferencing method as part of a full tracking system based on the JIPDA multi-target tracker. Other well known tracking methods such as Probabilistic Data Association (Bar-Shalom et al., 2011, p. 174) and Joint Probabilistic Data Association (Bar-Shalom et al., 2011, p. 387) are special cases of the JIPDA.

### 2.2 Measurement uncertainty

Accurate tuning of measurement uncertainty is important to yield state estimates in tracking that are statistically consistent. Setting the noise too small relative to the real value can yield a jumpy, measurement weighted state estimate. If the noise is too large the filter will be slow to respond to measurements, relying more on model predictions. We define the measurement uncertainty in pixel coordinates in the camera frame as this is where the detection system operates. This noise covariance matrix, $\Sigma^P$, is then converted to Cartesian world coordinates, $\Sigma^W$, based on the measurements or predictions according to

$$\Sigma^W = \mathrm{J}\Sigma^P\mathrm{J}^\mathrm{T} \tag{7}$$

where J is the Jacobian of (6) with respect to pixel position.

## 3. IMAGE DETECTION AND PROCESSING

Image data requires extensive processing to extract accurate detections. In this work we use a single shot detector (SSD) based deep learning method for object detection to extract bounding boxes which are then refined to provide more accurate detections.

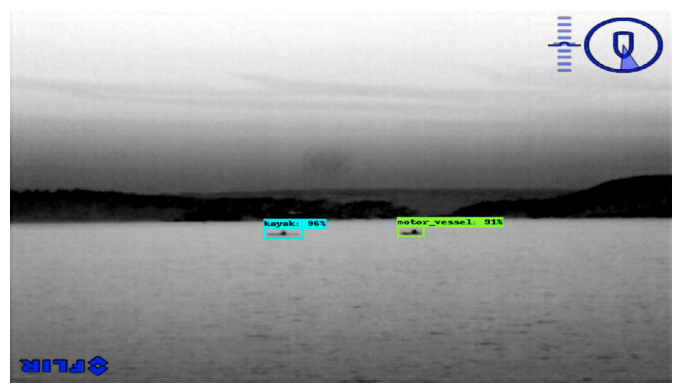


Fig. 1. IR detection output

### 3.1 SSD image data detector

Introduced in (Liu et al., 2015), the SSD functions by dividing an image into a grid consisting of a fixed amount of pre-computed bounding boxes. SSD learns these bounding boxes as part of the training process, known as MultiBox (Erhan et al., 2013). Regression is then employed to match these boxes to the actual objects within the image. This

allows SSD to combine both speed and accuracy. Deep learning based detectors such as SSD have been successfully used for object detection in maritime environments in (Helgesen et al., 2019; Schöller et al., 2019).

A Mobilenet v2, (Sandler et al., 2018), network pretrained on the COCO dataset, (Lin et al., 2014), was used as a base for the detector. A custom dataset consisting of 2035 images for each camera was labelled based on data recorded in 2017 at the same location. Using these images two separate detectors were trained using transfer learning, one for each camera type. Fig. 1 illustrates detector output.

### 3.2 Sobel operator

The Sobel operator, or Sobel filter, is an edge detection method for digital images. It is used to approximate the gradient of an image by convolving two $3 \times 3$ kernels with the image to find the horizontal and vertical gradients. These gradients, $\mathbf{G}_x$ and $\mathbf{G}_y$, are computed as

$$\mathbf{G}_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * \mathbf{I} \tag{8}$$

$$\mathbf{G}_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * \mathbf{I} \tag{9}$$

where $\mathbf{I}$ is an image.

### 3.3 Hough transform

The Hough transform, (Hough V, 1962), is a widely used method to detect geometric features in image data. The basic idea behind the Hough transform is that a line in image space has a corresponding point in the parameter space describing a line. Vice versa, a point in image space will result in a line in parameter space. The parametrization used today is due to (Duda and Hart, 1972) and is given by

$$\rho = x \cos\theta + y \sin\theta \tag{10}$$

where $[x, y]$ represents a point in an image and $[\rho, \theta]$ a point in parameter space. The actual line detection is done using a 2-dimensional accumulator array where each cell corresponds to a certain pair of parameters, $[\rho, \theta]$. If a line is detected in the neighbourhood of a pixel, the parameters of this line are found and the corresponding accumulator cell is incremented by one. Once completed for all pixels, the cells with the highest numbers will contain the most likely lines. A gradient image obtained from applying the Sobel operater is shown in Fig. 2. The points in the Hough plot, Fig. 3, with the largest number of intersecting lines corresponds to the cells in the accumulator array with the highest number.

## 4. PERFORMANCE METRICS

This section presents the metrics used to evaluate the various sensors and sensor combinations, both for track management and track accuracy. The MATLAB Sensor Fusion Toolbox was used to implement some of these metrics.

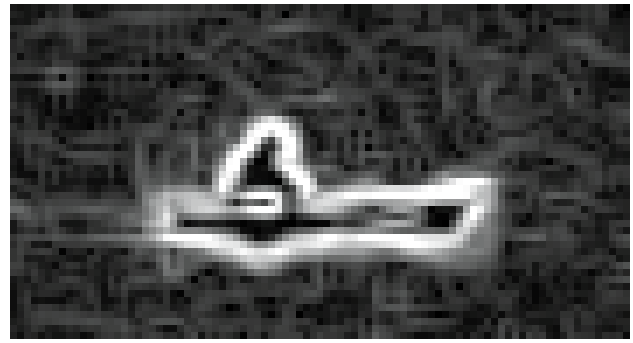

Fig. 2. Sobel gradient image of a detected boat

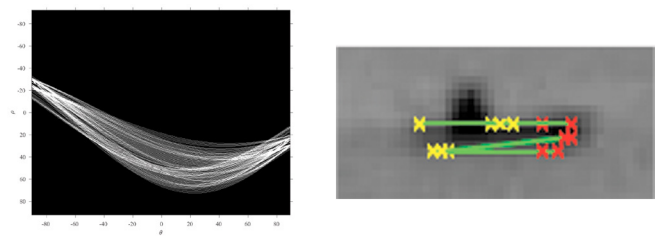

Fig. 3. Hough transform and the resulting Hough lines of Fig. 2

### 4.1 Track-truth assignment

The track-truth assignment determines whether a track originates from a valid target or from clutter. For every track the track-truth Euclidean distance is calculated to all current truths, if this distance is below a set threshold, 15m, the association is valid. The threshold was set based on smart phone GPS accuracy.

### 4.2 Track management metrics

Track management plays a vital part in the performance of a tracking system. Good track management can provide better track initialization, reduce the effect of false tracks and eliminate potential redundancies. This section presents a number of metrics designed to evaluate the track management performance of the tracking system.

- ***Establishment length*** evaluates the time in seconds required to establish a valid track-truth association measured from the start of a dataset. This serves as an estimate of how many seconds is required to establish a track once a target enters the surveillance region.
- ***False tracks*** are tracks not associated with a truth, originating from clutter and false detections. In this evaluation a false track is defined as a track that was never associated with a truth during its lifetime. This metric is reported as the average number of false tracks and their duration per dataset.
- ***Truth breaks*** occur when a truth becomes unassociated with a track, either due to track death or the track has associated with another truth. This metric is reported as the average track break time per dataset.

### 4.3 Track accuracy

Another area of key interest in evaluating tracking performance is the accuracy of the tracking results. Good track

management can be of little consequence if the resulting accuracy of the tracks are poor. Safe, autonomous maneuvering requires an accurate estimate of the current world state to avoid potential collisions with other object's. This section presents metrics designed to evaluate the accuracy of the tracking result decoupled from track management.

- ***Position accuracy*** is evaluated according to root-mean-square error (RMSE). RMSE is calculated for a single target-track pair according to

$$\text{PosRMSE} = \sqrt{\frac{\sum_{i=1}^{k}(\hat{\mathbf{x}}_i - \mathbf{x}_i)^2}{k}}, \tag{11}$$

  where $k$ is the total number of updates, $\hat{\mathbf{x}}_i$ and $\mathbf{x}_i$ the track and truth position. Position RMSE is calculated per target across all datasets.
- ***Divergence*** occurs when the Euclidean distance between a track-truth assignment exceeds the assignment threshold of 15m, that is

$$||\mathbf{x}_k^i - \hat{\mathbf{x}}_k^j|| > 15\text{m}. \tag{12}$$

Track deviation is a significant concern with regards to track breaks, large deviations can lead to valid measurements outside the validation gate of the track, increasing the probability of track death.

### *4.4 Filter consistency*

Since experimental data are used, filter consistency is evaluated with the Average Normalized Innovation Squared (ANIS). For a single target Kalman filter we have

$$\text{ANIS} = \frac{1}{N}\sum_{k=1}^{N} \mathcal{V}_k^T \mathbf{S}_k^{-1} \mathcal{V}_k, \tag{13}$$

where $\mathcal{V}_k$ is the innovation at time $k$ and $\mathbf{S}_k$ the innovation covariance. In a JPDA or JIPDA tracker multiple weighted Kalman filter updates can be used to update track states. In these cases the Normalized Innovation Squared (NIS) calculation for target $t$ is weighted according to the marginal association probabilities:

$$\text{NIS}_k^t = \frac{\sum_{j=1}^{m_k} \beta_k^{t,j} (\mathcal{V}_k^{t,j})^T (\mathbf{S}_k^t)^{-1} \mathcal{V}_k^{t,j}}{\sum_{j=1}^{m_k} \beta_k^{t,j}}, \tag{14}$$

where $\beta_k^{t,j}$ is the marginal association probability of track $t$ with measurement $j$ and $m_k$ the number of measurements. This metric is calculated for all tracks, valid or false, across all time steps and averaged to produce the reported ANIS metric. More information about the association probabilities can be found in (Musicki and Evans, 2004).

## 5. SENSORS AND EXPERIMENT SETUP

Electro-optical sensor data were provided by an AXIS P5514-E camera at a resolution of 1280 × 720 pixels. Infrared sensor data came from a FLIR M232 camera using a 320 × 240 VOx microbolometer sensor sensitive to long-wave infrared radiation. Both cameras were set to near identical fields of view, 24°, and sampled at 1Hz. The radar benchmark comes from a SIMRAD Broadband 4G radar.

### *5.1 Radar pipeline*

The radar used in this work contains a built-in detection system. These detections are presented in the form of spokes containing resolution cells corresponding to certain ranges and azimuth angles. Each cell contains a binary value representing whether a target is present or not in the range and azimuth covered by this cell. These resolution cells are converted into a 2D point cloud which is clustered to provide a single detection for each target. An in-depth exploration of this radar pipeline is available in (Wilthil et al., 2017).

The performance evaluation is based on several datasets recorded outside Oslo, Norway, at the DNV GL headquarters. Data were recorded using all sensors at both day and nighttime. Cameras were mounted to a mobile sensor rig provided by DNV GL, set at a fixed position on land at an elevation of 3 meters. In Fig. 4 the experiment area is shown with an approximate FOV for the cameras overlaid.

Fig. 4. Experiment area with camera FOV

Two reference targets were used in the evaluation, both recording a GPS ground truth using Android smart phones. Shown in Fig. 5, the reference targets include a small aluminium boat propelled by low-power electric motors and a kayak fitted with a radar reflector. Dataset weather conditions are given in Table 1. Some boating activity was present in the daytime dataset and a single non-reference target in the nighttime dataset. Any tracks resulting from these targets are assumed to be false by the evaluation system due to a lack of ground truths.

Fig. 5. Reference targets

Table 1. Testing conditions

| | **Dataset 1** | **Dataset 2** |
|---|---|---|
| **Light[LUX]** | 24k-28k | 0-7 |
| **Rain[mm]** | 0 | 0 |
| **Douglas Sea state** | 1-2 | 1 |

## 6. RESULTS

Using the datasets and metrics described previously the tracking performance of the georeferenced cameras, both individually and in fusion, are evaluated against a radar benchmark. Raw performance data can be found in Tables 2 and 3 as well as Fig. 8. A visualization of the tracking process at 200m range can be seen in Figs. 6 and 7.

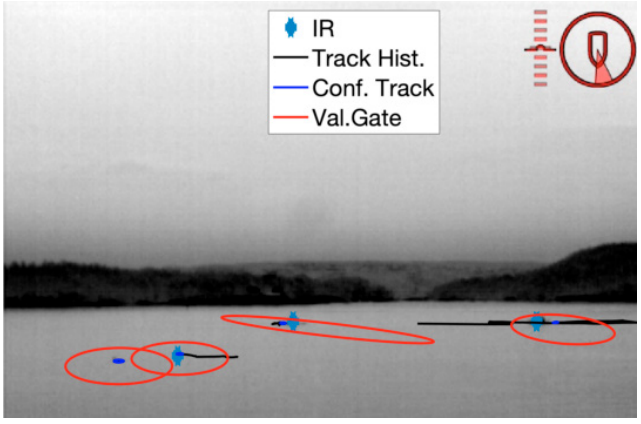


Fig. 6. IR tracking at 200m range

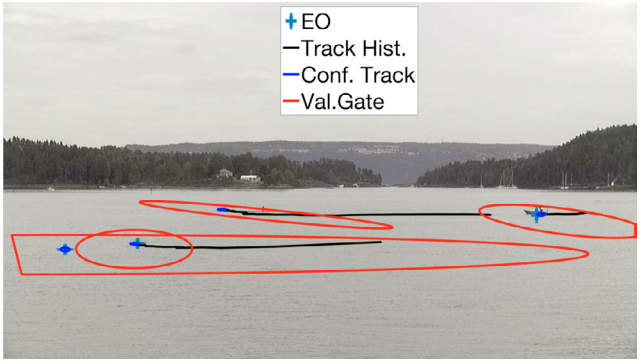


Fig. 7. EO tracking at 200m range

### 6.1 Electro-optical camera performance

In terms of track management the EO camera shows promising daytime performance compared to the radar. Tracks are established significantly quicker (Table 3) and could be improved further by sampling at the cameras native framerate. Track break times are also considerably lower although the number of breaks is an order of magnitude higher. False tracks (Table 2) are a problem, possibly due to the detectors tendency to detect sea birds as valid targets. Nighttime performance is significantly reduced due to a lack of illumination resulting in large amounts of sensor noise.

Track accuracy is reduced compared to the radar, both in terms of divergence time (Table 3) and in terms of RMS position error, Fig. 8. At certain ranges the camera does match the position accuracy of the radar but this is highly dependent on finding the correct pixel corresponding to the intersection between the vessel and the ocean. Due to the cameras low elevation and the long distances to targets minor pixel errors can result in position estimates off by tens of metres.

Table 2. Track metrics

| Sensors | False Tracks | False Track Length | ANIS |
|---|---|---|---|
| **R** | 20 | 262.10s | 2.06 |
| **IR** | 113 | 482.99s | 2.64 |
| **EO** | 87 | 564.67s | 2.14 |
| **IR,EO** | 223 | 376.53s | 2.63 |

### 6.2 Infrared camera performance

Compared to EO performance the IR camera is slightly slower in establishing tracks, possibly due to the lower resolution resulting in greater effects of detection inaccuracies. Both the number of track breaks as well as the track break time is roughly doubled compared to the EO camera. The nighttime performance of the IR camera is however much better than the EO camera due to the nature of the sensor with near identical performance regardless of lighting conditions. Similarly to the EO camera the IR camera yields better performance across several metrics compared to the radar. False tracks still remain a problem, though slightly less than with the EO camera.

For track accuracy the IR camera usually tracks the EO camera closely in terms of RMS position error, Fig. 8. Track divergence time is nearly doubled, increasing the likelihood of premature track deaths. The radar benchmark is still significantly better than both cameras in terms of divergences.

Table 3. Track Metrics, parantheses show daytime performance

| Sensors | Est. T. | Breaks | Break T. | Div. T. |
|---|---|---|---|---|
| **R** | 14.96s | 28 | 393.09s | 17.63s |
| **IR** | 5.14s (3.21s) | 341 | 307.18s | 196.73s |
| **EO** | 41.85s (1.41s) | 170 | 123.25s | 103.47s |
| **IR,EO** | 4.51s (0.57s) | 450 | 273.70s | 211.69s |

### 6.3 Sensor fusion performance

The effects of sensor fusion can result in significant advantages for robustness and redundancy (Helgesen et al., 2019). For track management the effects of sensor fusion are both positive and negative. Tracks are established quicker than with only EO. However the number of track breaks is greater than any of the cameras, although the break time is lower than the IR camera. A minor improvement in long range accuracy can be observed in Fig. 8. In certain cases performance was worse than individual cameras, possibly due to the lack of multi-sensor tuning.

## 7. CONCLUSION

A method for georeferencing imaging data from monocular cameras without horizon detection was presented in this paper and applied to maritime target tracking. Constraining objects to a flat plane modelling the ocean surface allowed position estimation using only a single camera. A reduction in track breaks was observed compared to the radar benchmark, possibly due to differences

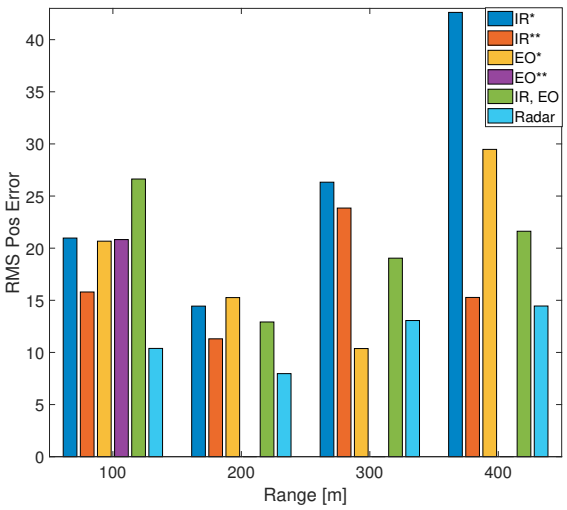


Fig. 8. Position RMSE. Asterisk signifies daytime only, double asterisk nighttime only.

in sensor resolution. The lower resolution of the radar was observed to cause merged measurements for targets at close ranges degrading tracking performance. We also observed reduced position accuracy compared to the radar benchmark. Higher camera resolutions and enhanced detection accuracy is expected to improve this along with further work on camera calibration. Planned future work includes application to vessel-mounted sensors where the sensor platform is in motion as well as integration into a heterogeneous multi-sensor fusion system with active and passive sensors.

## ACKNOWLEDGEMENTS

This work was made possible by DNV GL which provided a sensor rig complete with sensors.

## REFERENCES

Bar-Shalom, Y., Willett, P., and Tian, X. (2011). *Tracking and Data Fusion: A Handbook of Algorithms*. YBS Publishing.

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

Duda, R.O. and Hart, P.E. (1972). Use of the hough transformation to detect lines and curves in pictures. *Commununications of the ACM*, 15(1), 11–15.

Erhan, D., Szegedy, C., Toshev, A., and Anguelov, D. (2013). Scalable object detection using deep neural networks. *CoRR*, abs/1312.2249.

Fefilatyev, S., Goldgof, D., and Lembke, C. (2010). Tracking ships from fast moving camera through image registration. In *2010 20th International Conference on Pattern Recognition*, 3500–3503.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2980–2988.

Helgesen, H.H., Leira, F.S., Fossen, T.I., and Johansen, T.A. (2018). Tracking of ocean surface objects from unmanned aerial vehicles with a pan/tilt unit using a thermal camera. *Journal of Intelligent & Robotic Systems*, 91(3), 775–793.

Helgesen, Ø.K., Brekke, E., Helgesen, H., and Engelhardtsen, Ø. (2019). Sensor combinations in heterogeneous multi-sensor fusion for maritime target tracking. In *2019 22nd International Conference on Information Fusion (FUSION 2019)*. Ottawa, Canada.

Hough V, P.C. (1962). Method and means for recognizing complex patterns. U.S. Patent 30696541962.

Lin, T., Maire, M., Belongie, S.J., Bourdev, L.D., Girshick, R.B., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C.L. (2014). Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S.E., Fu, C., and Berg, A.C. (2015). SSD: single shot multibox detector. *CoRR*.

Musicki, D. and Evans, R. (2004). Joint integrated probabilistic data association: JIPDA. *IEEE transactions on Aerospace and Electronic Systems*, 40(3), 1093–1099.

Park, J., Kim, J., and Son, N. (2015). Passive target tracking of marine traffic ships using onboard monocular camera for unmanned surface vessel. *Electronics Letters*, 51(13), 987–989.

Prasad, D.K., Rajan, D., Rachmawati, L., Rajabally, E., and Quek, C. (2017). Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 18(8), 1993–2016.

Sandler, M., Howard, A.G., Zhu, M., Zhmoginov, A., and Chen, L. (2018). Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. *CoRR*, abs/1801.04381.

Schöller, F., Plenge-Feidenhans'l, M., Stets, J., and Blanke, M. (2019). Assessing deep-learning methods for object detection at sea from LWIR images. *12th IFAC Conference on Control Applications in Marine Systems, Robotics, and Vehicles (IFAC CAMS)*.

Szpak, Z.L. and Tapamo, J.R. (2011). Maritime surveillance: Tracking ships inside a dynamic background using a fast level-set. *Expert Systems with Applications*, 38(6), 6669 – 6680.

Wilthil, E.F., Flåten, A.L., and Brekke, E.F. (2017). A Target Tracking System for ASV Collision Avoidance Based on the PDAF. In T.I. Fossen, K.Y. Pettersen, and H. Nijmeijer (eds.), *Sensing and Control for Autonomous Vehicles: Applications to Land, Water and Air Vehicles*, 269–288. Springer International Publishing, Cham.

Woo, J. and Kim, N. (2016). Vision-based target motion analysis and collision avoidance of unmanned surface vehicles. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, 230(4), 566–578.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330–1334.