

# Sensor Combinations in Heterogeneous Multi-sensor Fusion for Maritime Target Tracking

Øystein Kaarstad Helgesen\*, Edmund Førland Brekke\*, Håkon Hagen Helgesen\* and Øystein Engelhardtson†

Department of Engineering Cybernetics, Norwegian University of Science and Technology (NTNU)\*

{oystein.k.helgesen, edmund.brekke, hakon.helgesen}@ntnu.no

DNV GL Norway†

oystein.engelhardtson@dnvgl.com

**Abstract**—Safe navigation for autonomous surface vehicles requires a robust and reliable tracking system that maintains and estimates position and velocity of other vessels. This paper demonstrates a measurement level sensor fusion system for tracking in a maritime environment using lidar, radar, electro-optical and infrared cameras. The backbone of the system is a multi-sensor version of the Joint Integrated Probabilistic Data Association (JIPDA) with both existence and visibility probabilities. Using reference targets equipped with GPS receivers, the performance of different sensors and sensor combinations are evaluated for autonomous surface vehicles (ASVs). Several interesting observations are made, among them that passive sensors can help resolve merged measurements issues in radar tracking, and that the choice between radar and lidar may boil down to a trade-off between fast track initiation and large numbers of false tracks.

**Index Terms**—Sensor fusion, unmanned surface vehicle, target tracking

## I. INTRODUCTION

It is generally recognized that automotive and maritime collision avoidance (COLAV) systems should utilize multiple sensors, as opposed to a single sensor, for improved robustness and reliability. The combination of heterogeneous sensors is particularly enticing, and the different sensors can complement each other and thus mitigate each others' weak spots. Nevertheless, a systematic empirical investigation of different sensor configurations is missing in the scientific literature.

Traditionally the key maritime sensor for navigation has been the radar augmented by human look-outs. The radar, while reliable, has challenges in certain situations and with certain target types. In these situations a human operated ship can rely on the look-outs to provide safe navigation. Autonomous vessels will not have this option, requiring a suite of several different sensors to move safely in different environments.

Several articles have been published in the field of target tracking and sensor fusion in recent years. [1] simulated a dual platform multi-sensor system for maritime border control. The

landbased platform was equipped with an infrared camera and Automatic Identification System (AIS) while an airborne platform was equipped with an infrared camera and radar. In [2], a tracking system based on probabilistic data association was used as part of a maritime COLAV system. [3] fused radar and camera data to perform obstacle detection for unmanned surface vehicles. They found that sensor fusion significantly increased obstacle tracking performance. [4] demonstrated a radar tracking system based on the PDAF tested on real data gathered in Trondheimsfjorden. Outside the maritime domain several papers have addressed heterogeneous sensor fusion for cars and land robots [5], [6].

This work is the first to demonstrate, evaluate, and describe in detail a sensor fusion system for maritime tracking using lidar, radar, infra-red and electro-optical cameras. This suite of sensors is relevant for autonomous operations both at land and sea. A Markov-chain two implementation of the JIPDA [7] method for multi-target tracking was used to test the various sensors and sensor combinations on a large, real-world dataset recorded using all four sensor types. This evaluation is based on a range of metrics covering both track management and accuracy [8].

This paper is organized as follows. Section II details the detection system used to extract measurements from sensor data. In section III the multisensor JIPDA tracking method is introduced. Section IV describes the metrics used to evaluate tracking performance. Section V introduces the sensors used in this work as well as the experimental setup. Results are given in section VI while section VII concludes this work.

## II. DETECTION SYSTEM

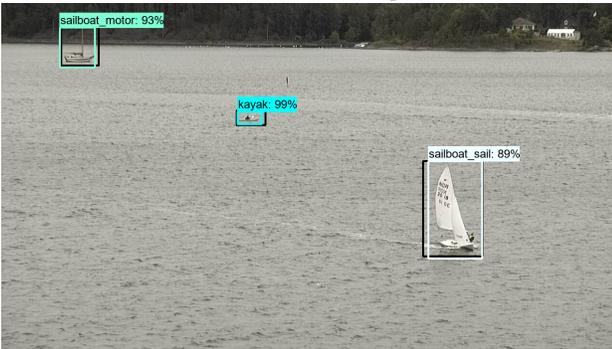
The sensor suite consists of four heterogeneous sensors, both active and passive. Data from a high-performance, long range lidar is mated to a clustering algorithm provided by the NTNU Revolve project [9] for detection. Sensor data are provided in the form of a high-resolution, 3D point cloud. Measurements are generated by clustering points based on geometric distance, using the cluster centroid as the target position since extended object tracking is outside the scope of this paper.

This work was supported by the Research Council of Norway (NFR) through the projects 223254 and 244116/O70. The professorship of Edmund Brekke is funded by DNV GL

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 [4].



(a) IR detection output



(b) EO detection output

Fig. 1: Visual detection output

### A. SSD image data detector

For object detection in both the infrared and visual spectrum the Single Shot Detector (SSD) [10] architecture is used. An image is divided 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 [11]. Regression is then employed to match these boxes to the actual objects within the image. This allows SSD to combine both speed and accuracy. SSD has e.g. been used for object detection in maritime environments in [12].

Hardware constraints made it impossible to run SSD detectors based on complex network architectures such as VGG 16 [13] in near real-time. Therefore, MobileNet 2 [14], which is a computationally cheaper neural net is used in this work to meet real-time requirements.

A Mobilenet v2 [14] based SSD detector pretrained on the COCO dataset [15] was used as a base for the detector. A custom dataset was labelled based on data recorded in 2017 at the same location. A combined total of 2035 images were manually labelled for both camera types with objects split into four categories inspired by COLREGs. These categories are kayaks, motor vessels and sailboats with and without sails. Using these images two separate detectors were trained using transfer learning, one for each camera type. Fig. 1 illustrates detector output for both cameras.

### III. JIPDA

The joint probabilistic data association method (JPDA) [16] is a multi-target extension of the probabilistic data association filter, PDAF [17]. This method accounts for uncertain data association by updating tracks according to a weighted sum of all observations within a validation gate. Each observation is weighted by their association probability, allowing a single observations to influence multiple tracks. An extension of JPDA accounting for target existence, and in some cases visibility, also exists in the form of the JIPDA [7] method. This work implements a Markov-chain two JIPDA which accounts for both target existence and visibility. It differs from the Markov Chain 2 JIPDA of [7] by decoupling the estimation of existence and visibility, leading to two Markov chains. [18] investigates the performance differences between the Markov chain one and two version of the single target JIPDA (IPDA).

For any target the target state is given by  $x = (y, v)$  where  $y$  is a continuous state vector and  $v$  a discrete component modelling visibility. The probability of detecting a target is then given by

$$P_D(x) = P_D(y, v) = \begin{cases} P_D & \text{if } v = O \\ 0 & \text{if } v = M \end{cases} \quad (1)$$

where  $v = O$  means the target is observable,  $v = M$  that the target is non-observable. The conventional single target likelihood modelling the relationship between a true detection,  $z$ , and target state is written as  $f_z(z|x)$ . This is only defined if the target is in a visible state,  $v = O$ . False alarms are modelled as a Poisson point process with intensity  $\lambda(z)$ .

#### A. Predictions

The existence prediction is given by

$$\begin{bmatrix} \epsilon^t \\ 1 - \epsilon^t \end{bmatrix} = \begin{bmatrix} p_{11}^\epsilon & p_{12}^\epsilon \\ p_{21}^\epsilon & p_{22}^\epsilon \end{bmatrix} \begin{bmatrix} \hat{\epsilon}_{k-1}^t \\ 1 - \hat{\epsilon}_{k-1}^t \end{bmatrix}, \quad (2)$$

where  $\epsilon^t$  denotes the predicted existence probability for target  $t$  at time  $k$ .  $\hat{\epsilon}^t$  is the posterior existence probability at time  $k$  for target  $t$ . Target visibility is similarly given by

$$\begin{bmatrix} \eta^t \\ 1 - \eta^t \end{bmatrix} = \begin{bmatrix} p_{11}^\eta & p_{12}^\eta \\ p_{21}^\eta & p_{22}^\eta \end{bmatrix} \begin{bmatrix} \hat{\eta}_{k-1}^t \\ 1 - \hat{\eta}_{k-1}^t \end{bmatrix}, \quad (3)$$

where  $\eta^t$  is the predicted visibility probability and  $\hat{\eta}_{k-1}^t$  the posterior visibility probability.

Prediction is done using the Chapman-Kolmogorov equation based on the prior distribution,  $\hat{f}(y_{k-1}^t)$ , as well as the kinematic model,  $f_y(y^t|\hat{y}_{k-1}^t)$ .

$$f^t(y) = \int \hat{f}_y(y^t|\hat{y}_{k-1}^t)\hat{f}_{k-1}^t(y_{k-1}^t)dy_{k-1}^t \quad (4)$$

For a Gaussian linear target model, such as the constant velocity model used in this work, and a Gaussian prior, the expected value and covariance will be given by the Kalman filter prediction.

### B. Updates

The posterior distribution of the JIPDA is given by a multi-Bernoulli mixture over association hypotheses. An association hypotheses, or association event, can be represented as a mapping

$$\theta : \{1, \dots, n\} \rightarrow \{0, 1, \dots, m\}, \quad (5)$$

where  $\{1, \dots, n\}$  is tracks and  $\{0, 1, \dots, m\}$  measurements. A mapping from any track to 0 implies no measurements originated from that target in the current hypothesis. For any event the tracks can be split into tracks on detected targets,  $D(\theta)$ , and tracks on undetected targets,  $M(\theta)$ . For any track on detected targets the following are defined

$$l^{t,\theta(t)} = \int f_z(z^{\theta(t)}|y)f(y)dy \quad (6)$$

$$\hat{f}^{t\theta(t)}(y) \propto f_z(z^{\theta(t)}|y)f(y) \quad (7)$$

where  $l^{t,\theta(t)}$  is the likelihood of the association of track  $t$  with a measurement given by the event  $\theta(t)$ .  $\hat{f}^{tj}(y)$  represents the posterior state density of track  $t$  given an association event. The probability of any event can be found as

$$P\{\theta\} \propto \prod_{j \notin \text{Im}(\theta)} \lambda(z^j) \prod_{t \in D(\theta)} \epsilon^t l^{t,\theta(t)} P_D \eta^t \cdot \prod_{t \in M(\theta)} (1 - \epsilon^t + \epsilon^t (1 - P_D) \eta^t). \quad (8)$$

The probabilities of a measurement  $z^j$  originating from target  $t$ , the marginal association probabilities, are given by

$$p^{tj} = \sum_{\theta \text{ s.t. } \theta(t)=j} P\{\theta\}. \quad (9)$$

The posterior existence probability conditional on no measurement association for target  $t$ , denoted by  $\hat{\epsilon}_k^{t,0}$ , is

$$\hat{\epsilon}_k^{t,0} = \frac{1 - P_D \eta^t}{1 - \epsilon^t P_D \eta^t} \epsilon^t, \quad (10)$$

where  $P_D$  is the detection probability. For visibility probability we have

$$\hat{\eta}_k^{t,0} = \frac{\eta^t (1 - P_D^t)}{1 - P_D^t \eta^t}. \quad (11)$$

For any track,  $t$ , the marginal existence and observability probabilities as well as the kinematic probability density function will be given by

$$\hat{\epsilon}^t = p^{t0} \hat{\epsilon}^{t0} + \sum_{j=1}^m p^{tj} \quad (12)$$

$$\hat{\eta}^t = \frac{1}{\hat{\epsilon}^t} \left( p^{t0} \hat{\epsilon}^{t0} \hat{\eta}^{t0} + \sum_{j=1}^m p^{tj} \right) \quad (13)$$

$$\hat{f}^t(y) \propto p^{t0} \hat{\epsilon}^{t0} f^t(y) + \sum_{j=1}^m p^{tj} \hat{f}^{tj}(y). \quad (14)$$

A complete derivation is available in [19, p. 53-60].

## IV. TRACKING 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.

### A. Track-truth assignment

The track-truth assignment determines whether a track is associated with a truth or originates from clutter. At every time-step any track not currently associated with a truth and any divergent track take part in the assignment process. Divergence is described in section IV-C2. 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. This threshold was set relatively large to compensate for potential GPS inaccuracies. If several track-truth associations meet this threshold the association minimizing track-truth distance is selected. Equation 15 describes this assignment as an optimization problem where  $\mathbf{x}_k^i$  and  $\hat{\mathbf{x}}_k^j$  are respectively the positions of the  $i$ th truth and  $j$ th track at time-step  $k$ . Each track can only be associated with a single truth, however a truth can be associated with multiple tracks.

$$\min_i \|\mathbf{x}_k^i - \hat{\mathbf{x}}_k^j\| \quad \text{s.t.} \quad \|\mathbf{x}_k^i - \hat{\mathbf{x}}_k^j\| \leq 15\text{m} \quad (15)$$

### B. Track management metrics

Track management plays a vital part in the performance of a tracking system. Good track management can provide better track initializations, 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.

1) *Establishment length*: The establishment length metric evaluates the number of time-steps required to establish a valid track-truth association measured from the start of the dataset. This metric is calculated for each truth across all the evaluated datasets and then summed. As a final step this sum is normalized by the total number of samples to compensate for the varying number of samples across sensors/sensor combinations. An example calculation for a 210 sample dataset with two targets present follows. Target one has a track established after 15 samples, yielding an establishment length

of 15. Target two has a track established after 20 samples, thus the establishment length is 20. To compute the establishment length metric these are summed and then normalized by the total number of samples:

$$\text{Est.L.} = \frac{15 + 20}{210} = 0.167. \quad (16)$$

2) *False tracks*: A false track is a track 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 number, F.T., is reported as a sum across all datasets. In addition, the lengths of the false tracks are also recorded and reported as false track length, F.T.L. This number is the sum of the lengths of all the false tracks, normalized by the total number of time-steps for the given sensors.

3) *Truth breaks*: A truth break occurs when a truth becomes unassociated with a track, either due to track death or the track has associated with another truth. The number of breaks is summed across all datasets for every truth. In addition, the length of the breaks is also summed across all datasets and truths and then normalized according to the total number of updates.

### C. 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 objects. This section presents metrics designed to evaluate the accuracy of the tracking result decoupled from track management.

1) *Position accuracy*: Perhaps the most basic metric for track accuracy, the position accuracy is a measure of the difference between the actual position of a target and its corresponding track. Position accuracy is evaluated according to RMS error. RMS, or root-mean-square, error is calculated for a single target-track pair according to

$$\text{PosRMS} = \sqrt{\frac{\sum_{i=1}^k (\hat{\mathbf{x}}_i - \mathbf{x}_i)^2}{k}}, \quad (17)$$

where  $k$  is the total number of updates,  $\hat{\mathbf{x}}_i$  and  $\mathbf{x}_i$  the track and truth position. RMS position error is calculated for both targets across all datasets and is reported per truth. These values are then averaged to provide a metric influenced by all available datasets.

2) *Divergence*: Mentioned in several of the metrics above, track-truth divergence occurs when the Euclidean distance between a track-truth assignment exceeds the assignment threshold of 15m, that is

$$\|\mathbf{x}_k^i - \hat{\mathbf{x}}_k^j\| > 15\text{m}. \quad (18)$$

Divergence is reported as two separate metrics. The divergence count, Div, reports the number of times a track entered a divergent state across all datasets. A second metric known

as divergence length, DivL, reports the number of updates a track was in a divergent state summed across all datasets. This value is normalized by the total number of samples.

### D. Filter consistency

Since experimental data are used, filter consistency is evaluated with the Average Normalized Innovation Squared (ANIS). Strictly speaking both ANIS and ANEES are metrics of filter consistency, ANEES does however rely on the true target position. In experimental data, a reliable truth is not necessarily available, and ANEES can be misleading. Therefore, for real data, ANIS is often used instead [20].

For a single target Kalman filter ANIS is given by

$$\text{ANIS} = \frac{1}{N} \sum_{k=1}^N \mathcal{V}_k^T \mathbf{S}_k^{-1} \mathcal{V}_k, \quad (19)$$

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 NIS calculation for target  $t$  is weighted according to the marginal association probabilities:

$$\text{NIS}_k^t = \frac{\sum_{j=1}^{m_k} \beta_k^{t,j} (\mathcal{V}_k^{t,j})^T (\mathbf{S}_k^t)^{-1} \mathcal{V}_k^{t,j}}{\sum_{j=1}^{m_k} \beta_k^{t,j}}, \quad (20)$$

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.

## V. SENSORS AND EXPERIMENT SETUP

The NEPTEC Opal 3 lidar used in this work relies on rotating prisms to scan the scene using only a single laser beam and a pair of rotating prisms. Compared to most array-based lidars which use multiple laser beams scanning along fixed lines, this approach allows for greater spatial resolution. This can, however, come at the cost of lower temporal resolution due to scan pattern movements. The lidar used in this experiment has a specified range of 1000m and a field of view (FOV) of 120° [21].

Radar data are provided by a SIMRAD Broadband 4G radar [22]. This is a frequency modulated continuous wave radar with a built-in detection system. Radar beam width is 5.2° but can be adjusted. An AXIS P5514-E [23] camera provides electro-optical (EO) images. This camera is capable of pan, tilt and zoom operations and has a resolution of 1280 × 720 pixels. Sensor data is provided in the form of a MJPEG stream which is sampled at 1Hz. Infrared images are provided by a FLIR M232 camera [24]. It features a VOx microbolometer sensor with a resolution of 320 × 240 pixels and a FOV of 24°. The AXIS camera was zoomed to match this. Images are provided as an H.264 stream which is sampled at 1Hz.

The tracking performance of the various sensor combinations was evaluated on a dataset recorded on a sensor rig provided by DNV GL. Several datasets were recorded at DNV GL Høvik during the summer of 2018 using all sensors in various lighting conditions. The sensor rig was mounted on land. Fig. 2 shows the experiment area and an approximate field of view for the cameras.



Fig. 2: Experiment area, image provided by Google Maps.

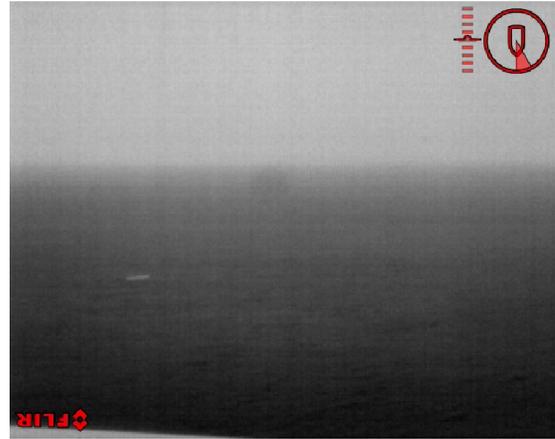
The datasets contain two reference objects, one small aluminium boat, target 1, and a kayak, target 2, in addition to general boating activity. The kayak was fitted with a radar reflector and can be seen in Fig. 4. Several maneuvers were performed at ranges from 100-500m with increments of 100m. This includes turns and crossing of objects in addition to straight line crossings of the experiment area. During recordings the sensor rig was positioned at the waterfront at an elevation of roughly 2m above sea level. This was done to simulate smaller vessels were sensors must be mounted closer to sea level. Data were recorded in two different lighting conditions, cloudy weather (24kLux) and at night (0-7 Lux), illustrated in Fig. 1 and Fig. 3. Weather conditions were clear in all scenarios with no precipitation and low wave heights. A minor amount of general boating activity was present during the recording of the cloudy dataset, these would be assumed to be false tracks in the implemented evaluation system due to the lack of ground truths. A single non-reference vessel was present during parts of the night dataset recording. A summary of testing conditions can be seen in TABLE I.

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

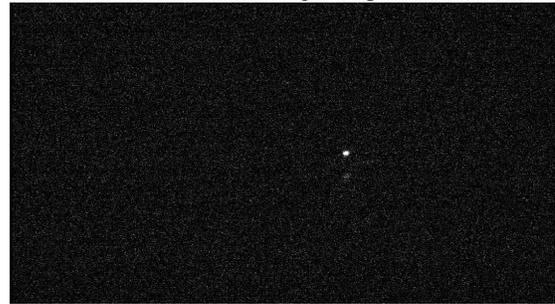
TABLE I: Testing conditions

## VI. RESULTS

In this section the results of the evaluation of the various sensors and sensor combinations are presented using the



(a) IR camera output, nighttime

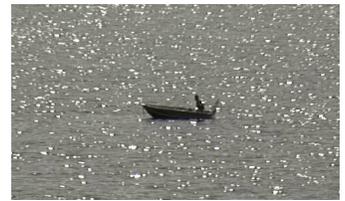


(b) EO camera output, nighttime

Fig. 3: Nighttime camera outputs.



(a) Kayak



(b) Small motor vessel

Fig. 4: Reference targets

metrics and datasets presented in sections IV and V. Results are presented for both active fusion and active-passive fusion, referring to the nature of the sensors involved.

### A. Active fusion

Active fusion refers to the fusion of measurements from active sensors, in this case lidar and radar. This section presents the tracking results of the active sensors, both individually and in fusion. These sensors are evaluated using the tracking metrics defined in section IV to provide an insight into the strengths and weaknesses of the sensors in a target tracking system.

1) *Track management*: In terms of track management, TABLE II, the lidar has some glaring weaknesses. The number of false tracks is an order of magnitude higher than the radar. Filtering out all detections at less than 100m, necessitating the removal of these datasets from the evaluation, reduces

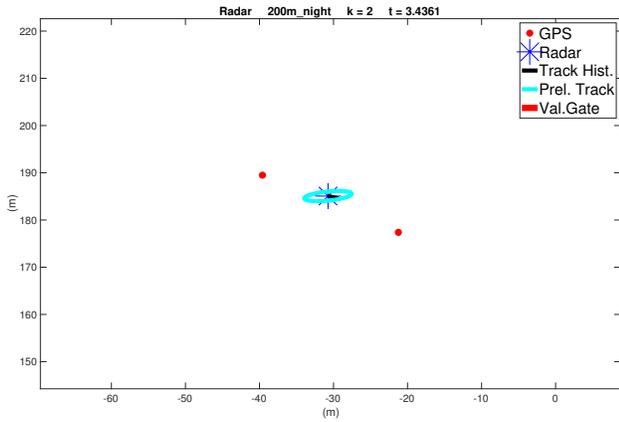


Fig. 5: Example of merged radar measurement

the number of false tracks significantly from 369 to only 20 which is a more acceptable number. This also suggests that improvements in the detector and elevation based filtering of the point cloud could result in a significant reduction in false tracks by eliminating most of the short range false detections.

In contrast, the radar provides 21 false tracks which is a significantly lower amount compared to the lidar. The radar did however yield establishment lengths 2.5 times greater than the lidar, suggesting that track initialization could be improved by sensor fusion. Readers should be aware that while the radar does contain a built-in filtration system that removes false echoes, it is unknown whether the lidar performs a similar operation.

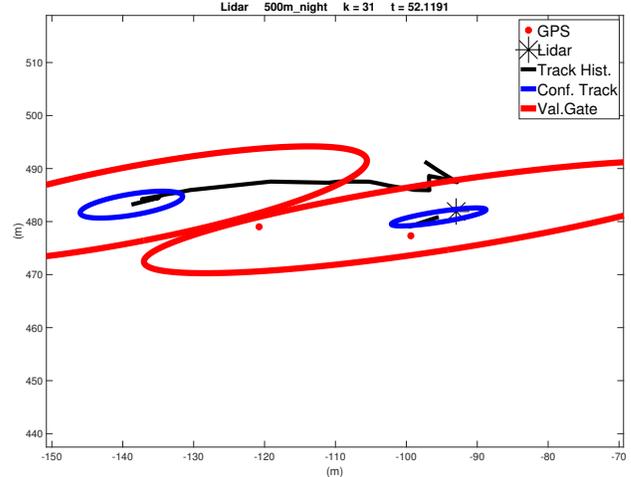
The track break metrics are slightly better for the radar. One possible explanation is the performance decrease the lidar experiences at longer ranges could cause valid tracks to die due to a series of missed detections. The radar would not suffer from this due to its excellent performance at all ranges.

Fusing radar and lidar measurements provide some notable improvements over both sensors. Track establishment is improved resulting in valid tracks being formed faster. The unreliability of lidar detections at longer ranges could cause an object to only be detected in one of two consequent measurement vectors, resulting in no track being formed. A combination of the reliability of the radar with the fast update rate of the lidar could be the cause of this increase. The number of track breaks does rise but without a significant increase in break length. Similarly the number of false tracks increases, however the total length of the false tracks is reduced.

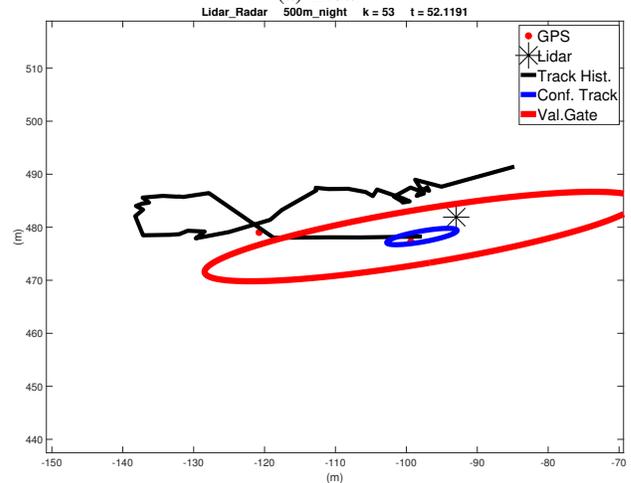
2) *Track accuracy*: In terms of track accuracy, TABLE III, the lidar shows it strengths over the radar. While the number of divergences is nearly identical the combined divergence length is an order of magnitude lower for the lidar. Manual inspection of the dataset revealed that the radar suffers from merged measurements when targets are close. This merged

measurement is positioned somewhere between the two targets, Fig. 5, and could cause both tracks to diverge for as long as the merged measurement persists. RMS position error is improved by roughly 3m for both targets by using the lidar over radar, again possibly due to merged measurements.

Fusing radar and lidar measurements provide some



(a) Lidar



(b) Lidar-Radar fusion

Fig. 6: Lidar and lidar-radar fusion at night, 500m range

marked improvements in track accuracy but is not without its drawbacks. Fig. 6 shows how merged measurements from the radar can result in only a single track for close targets. In addition, the number of divergences is increased but the total length is lower. If each sensor experiences divergences at different times one sensor could cause a divergence that the other sensor rapidly corrects. This implies that divergences persist over some time, suggesting they are not independent events. Combining lidar and radar improves RMS position error compared to both sensors, sensor fusion can thus yield a system better than its individual parts.

3) *Filter consistency*: Radar only results in ANIS very close to a 95% confidence interval. Minor tuning of the filter

parameters should push this into the accepted range of values. For both lidar and lidar-radar fusion ANIS is below the given confidence interval. This results in greater position fluctuations and could originate from several causes. Filter tuning could improve performance, but might come at the expense of radar performance. Another possible cause is the high amount of false tracks at shorter ranges. Point cloud density at longer ranges could also adversely affect ANIS.

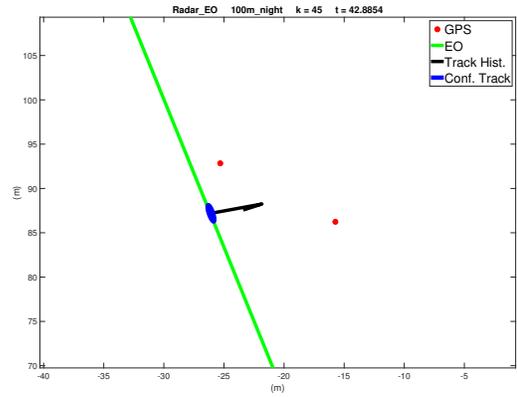
### B. Active-passive fusion

Adding passive sensors to a sensor fusion can give rise to many interesting effects. The passive detection system provides only a bearing measurement but was found to be very good at detection up to ranges of 400m. Fusing these bearing measurements with the active position measurements could thus result in improved performance. This section evaluates active-passive fusion using the same setup as the active fusion section.

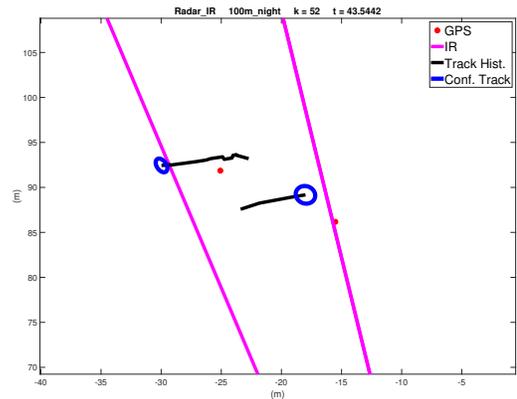
1) *Track management*: In terms of establishment length the combination of lidar, radar and electro-optical camera reigns supreme. The combinations of lidar, radar, IR and lidar, radar, EO, IR come close but are in the end unable to match the track establishment performance of a fusion of lidar, radar and EO. This result is somewhat surprising, the EO camera performs poorly in low-light conditions, Fig 7, which make up half of the datasets. This should suggest that fusing IR measurements would be better, which is not the case.

Two possible explanations present themselves. Either the EO cameras daylight performance is so much better compared to the IR camera that it outweighs the lack of low-light detections or the IR camera actually worsens track establishment when fused with radar and lidar measurements. Track establishment lengths for only radar and lidar, found in TABLE II, seem to support this. Track establishment lengths are lower when only lidar and radar is used compared to the addition of IR measurements. This degradation in performance is not observed for fusion of lidar with IR but is present for all the other fusions of a single active and passive sensor.

One area where active-passive fusion always results in better performance is track breaks. Adding a passive sensor reduces both the number of track breaks and their lengths regardless of which active and passive sensor is used. Passive fusion can allow the tracking system to keep valid tracks alive when no active measurements are available. Fusing lidar with IR and EO reduces break length by roughly 20% compared to lidar alone. Radar tracking experiences an even greater reduction of 30% when it is fused with IR. A degradation in position accuracy is expected in these cases, the passive sensors only measure a targets bearing, but the tracks remain alive and the position divergence was observed to be minor in most cases. This trade-off seems to be worthwhile in most cases. A valid and alive track can still be used for navigation purposes even when the position estimates are less accurate.



(a) Radar-EO fusion



(b) Radar-IR fusion

Fig. 7: Comparison of IR and EO fusion at night, 100m range

Based on TABLE II IR fusion seems to offer superior performance over EO fusion when considering only track breaks. Compared to EO fusion IR fusion offers lower track breaks and track lengths, likely due to its superior low-light detection performance. Adding EO measurements to a fusion already containing IR measurements can actually worsen performance when only fused with radar measurements. When lidar measurements are part of the fusion system adding EO measurements reduce break lengths but increase the number of breaks.

False tracks are again an issue for sensor combinations including the lidar. The high amount of false detections at shorter ranges massively increases the number of false tracks and the combined false track length. One trend observed across most sensor combinations is an increase in false tracks when more sensors are used. Every sensor brings with it a certain number of false detections. These detections seems to result in a compound effect adversely affecting the number of false tracks. Evaluating the fusion of all sensors but without the data recorded at a range of 100m reduces the number of

Sensors	Est.L.	Breaks	Break lengths	NumSamples
L	0.1511	48	0.7467	4501
R	0.3795	34	0.5956	361
L,R	0.1337	58	0.7649	4862
L,IR	0.1489	39	0.6147	5238
L,EO	0.1590	47	0.6922	5039
L,IR,EO	0.1584	44	0.5985	5776
L,R,IR	0.1390	38	0.6148	5599
L,R,EO	0.1300	53	0.7144	5400
L,R,IR,EO	0.1351	41	0.5982	6137
R,IR	0.4454	30	0.4180	1098
R,EO	0.4905	33	0.5273	899
R,IR,EO	0.4847	36	0.4230	1636
Sensors	False Tracks		False Track Length	
L	373		2.2731	
R	21		0.3241	
L,R	387		2.1421	
L,IR	396		2.0682	
L,EO	388		2.1478	
L,IR,EO	407		1.9778	
L,R,IR	409		1.9664	
L,R,EO	399		2.0352	
L,R,IR,EO	419		1.8882	
R,IR	60		0.1712	
R,EO	62		0.1591	
R,IR,EO	60		0.1351	

TABLE II: Track metrics

false tracks to 25 and the false track length to 0.2902, a very significant improvement. The number of track breaks is also halved along with a reduction of redundant tracks to 9. This highlights the importance of detector performance in target tracking. Introducing a range-dependent false alarm rate as was done with detection performance could potentially reduce the number of false tracks along with more advanced track initialization methods.

2) *Track accuracy*: Adding passive sensors to the sensor fusion system provides some interesting effects on track divergence. With a few exceptions both the number of divergences and their lengths have increased. Lidar alone had 30 divergences with a length of 0.0573. Fusing lidar with IR increased divergences to 31 with a length of 0.0878. A fusion of radar and EO is the only combination that actually improves performance across all divergence metrics, reducing divergences from 29 to 27 and their lengths from 0.1745 to 0.1035. Replacing EO with IR actually worsens performance even though both sensors provide similar performance. Active sensors, either lidar alone or in fusion with radar, remains the best choice for reducing divergence lengths. A likely explanation is that the addition of bearing measurements to the position estimates could provide longer divergences when targets are maneuvering. Bearings only tracking relies much more on model predictions, thus accuracy will suffer when the constant velocity assumption of the model is violated, e.g. during maneuvering.

The root mean square position error is, with a few exceptions, actually improved with the addition of a passive sensor. Both IR, EO and their combination slightly reduce the RMS position error when fused with lidar. The same effect is

Sensors	ANIS	Divergence lengths	RMS
L	0.5364	0.0573	[10.7; 8.4]
R	2.3886	0.1745	[13.6;11.2]
L,R	0.6012	0.0527	[7.5;7.9]
L,IR	0.5441	0.0878	[12.3; 8.3]
L,EO	0.7121	0.0929	[8.4;9.8]
L,IR,EO	0.6926	0.0966	[ 8.8;8.7]
L,R,IR	0.6190	0.1062	[14.9; 8.2]
L,R,EO	0.7749	0.1100	[8.2;8.9]
L,R,IR,EO	0.7562	0.1480	[9.5;9.3]
R,IR	1.6601	0.2031	[11.0;11.2]
R,EO	2.2429	0.1035	[9.8;8.8]
R,IR,EO	1.7980	0.1712	[9.2;10.8]

TABLE III: Track metrics. RMS is given per target

observed when fused with radar. Manual inspection of the datasets revealed that the lidar often struggled with reliably detecting targets at longer ranges. Consecutive measurements of a target could therefore arrive with some interval, potentially reducing accuracy. The detection reliability of the passive sensors could be the cause of this improvement, even though they only provide target bearings.

While this explanation could explain the improvements of lidar-based tracking, the radar was evaluated to be extremely reliable. The improvements to radar-based tracking must therefore have a different cause. As mentioned previously the radar did in some cases suffer from merged measurements. In effect, the radar would generate only a single, merged detection when two targets were sufficiently close to each other. This would undoubtedly reduce the accuracy of any position estimate based on this measurement, hence the increased RMS position error. No such effect was ever observed for any of the passive sensors. Accordingly, a fusion of radar with passive sensors might help mitigate the effects of merged measurements. It should be noted that most sensor combinations are very close in terms of RMS position error. Any variances between these sensor combinations could in practice be caused by GPS inaccuracies, not the superiority of a particular sensor or sensor combination.

3) *Filter consistency*: Similarly to active fusion the filter consistency metrics for active-passive fusion are better than the position accuracy metrics. Sensor combinations containing the lidar seem to result in ANIS values grouped around 0.6 while radar fusions are grouped around 2. A fusion of radar and EO or radar and IR results in ANIS values barely outside the confidence intervals, slightly improving the results of radar only tracking. Fusions containing the lidar experience far greater gaps to the confidence interval bounds. This suggests further filter tuning is required, especially for the lidar.

## VII. CONCLUSION

This work demonstrates a working implementation of measurement level sensor fusion for the purpose of detection and tracking of objects at sea. Four different types of active and passive sensors were used to evaluate sensor fusion: lidar, radar, infra-red camera and electro-optical camera. A detection

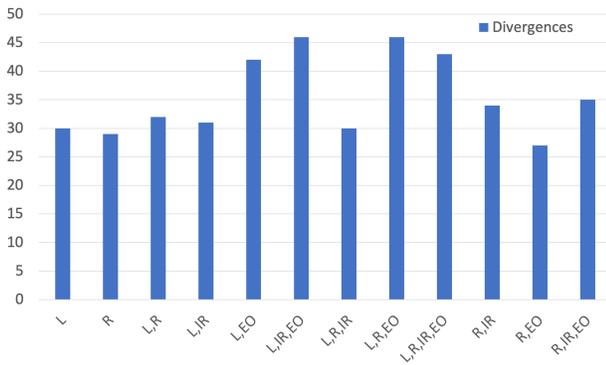


Fig. 8: Track divergence

system taking inputs of raw sensor data from all sensors and outputting vectors of measurements was implemented in the Robot Operating System using state-of-the-art methods. Using a JIPDA tracker the performance of sensors and sensor fusions was then evaluated from a tracking perspective.

Radar alone was very reliable for already initialized tracks but often struggled with track birth due to merged measurements. This also caused longer divergences between tracks and targets and a higher position error. Lidar provides better track initialization, but the low probability of detection at longer ranges caused tracks to die prematurely resulting in longer track breaks. A combination of lidar and radar yields lower establishment lengths, fewer divergences and a higher position accuracy.

Adding passive sensors can improve tracking results further. Fusing IR measurements did always provide some sort of positive effect. Reliable detections at longer ranges required radar, preferably in fusion with lidar to mitigate merged measurements. Adding the IR camera to this mix provided a significant improvement in the track break metrics and yielded low establishment lengths. A combination of radar, lidar and IR camera is therefore, in the author's opinion, the most robust sensor combination for this dataset among those evaluated in this paper. This conclusion is based on establishment lengths and track divergence in addition to manual inspection of tracking results. Future research includes testing other methods (e.g. PMBM, MHT) as well as sensors more similar in cost, this work pairs an expensive lidar with a low cost radar.

#### ACKNOWLEDGMENT

This work was made possible by DNV GL which provided a sensor rig complete with sensors. NEPTEC provided the lidar used in this paper.

#### REFERENCES

- [1] S. Giompapa, A. Farina, F. Gini, A. Graziano, and R. di Stefano, "Computer simulation of an integrated multi-sensor system for maritime border control," in *2007 IEEE Radar Conference*, pp. 308–313, April 2007.
- [2] M. Schuster, M. Blaich, and J. Reuter, "Collision Avoidance for Vessels using a Low-Cost Radar Sensor," in *19th IFAC World Congress*, vol. 47, pp. 9673 – 9678, 2014.
- [3] D. Hermann, R. Galeazzi, J. Andersen, and M. Blanke, "Smart Sensor Based Obstacle Detection for High-Speed Unmanned Surface Vehicle," in *10th IFAC Conference on Manoeuvring and Control of Marine Craft MCMC 2015*, vol. 48, pp. 190 – 197, 2015.
- [4] E. F. Wilthil, A. L. Flåten, and E. F. Brekke, "A target tracking system for ASV collision avoidance based on the PDAF," in *Sensing and Control for Autonomous Vehicles* (P. Fossen and Nijmeijer, eds.), vol. 474, pp. 269 – 288, Alesund, Norway: Springer, 2017.
- [5] M. Haberjahn and K. Kozempel, "Multi level fusion of competitive sensors for automotive environment perception," in *Proceedings of the 16th International Conference on Information Fusion*, pp. 397–403, July 2013.
- [6] H. M. Barbera, A. G. Skarmeta, M. Z. Izquierdo, and J. B. Blaya, "Neural networks for sonar and infrared sensors fusion," in *Proceedings of the Third International Conference on Information Fusion*, vol. 2, July 2000.
- [7] D. Musicki and R. Evans, "Joint integrated probabilistic data association: JIPDA," *IEEE transactions on Aerospace and Electronic Systems*, vol. 40, no. 3, pp. 1093–1099, 2004.
- [8] "Sensor Fusion and Tracking Toolbox - MATLAB & Simulink." <https://se.mathworks.com/products/sensor-fusion-and-tracking.html>. (Accessed on 01/31/2019).
- [9] "NTNU Revolve." <https://www.revolve.no/>. (Accessed on 12/05/2018).
- [10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg, "SSD: single shot multibox detector," *CoRR*, 2015.
- [11] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," *CoRR*, vol. abs/1312.2249, 2013.
- [12] S. V. Grini, "Object Detection in Maritime Environments," Master's thesis, Norwegian University of Science and Technology, 2019.
- [13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.
- [14] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation," *CoRR*, vol. abs/1801.04381, 2018.
- [15] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: common objects in context," *CoRR*, vol. abs/1405.0312, 2014.
- [16] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffe, "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, Dec 1980.
- [17] Y. Bar-Shalom and E. Tse, "Tracking in a cluttered environment with probabilistic data association," *Automatica*, vol. 11, no. 5, pp. 451 – 460, 1975.
- [18] E. F. Wilthil, Y. Bar-Shalom, P. Willett, and E. F. Brekke, "Estimation of Target Detectability for Maritime Target Tracking in the PDA Framework." Submitted to: 22nd International Conference on Information Fusion (FUSION), 2019.
- [19] Øystein Kaarstad Helgesen, "Sensor Fusion for Detection and Tracking of Maritime Vessels," Master's thesis, Norwegian University of Science and Technology, 2019.
- [20] Y. Bar-Shalom, T. Kirubarajan, and X.-R. Li, *Estimation with Applications to Tracking and Navigation*. New York, NY, USA: John Wiley & Sons, Inc., 2002.
- [21] "OPAL 3 P Series datasheet." [http://www.neptectechnologies.com/wp-content/uploads/2018/06/OPAL-3-P-Series\\_-conical-FOV-88-00201-01-06\\_2018-1.pdf](http://www.neptectechnologies.com/wp-content/uploads/2018/06/OPAL-3-P-Series_-conical-FOV-88-00201-01-06_2018-1.pdf). (Accessed on 10/19/2018).
- [22] "SIMRAD Broadband 4G Radar Brochure." <https://ww2.simrad-yachting.com/Root/Brochures/SimradYachting/English/Simrad%204G%20Brochure-AMER.pdf>. (Accessed on 10/19/2018).
- [23] "AXIS P5514-E PTZ network camera | Axis Communications." <https://www.axis.com/it/en/products/axis-p5514-e>. (Accessed on 10/12/2018).
- [24] "FLIR M232 Compact Pan/Tilt Marine Thermal Camera | FLIR Systems." <https://www.flir.com/products/m232/>. (Accessed on 10/12/2018).