Track-to-Track AIS / Radar Association and Uncertainty Estimation by Coherent Point Drift

Valérian MANGÉ
Hensoldt Nexeya France[1]
TéSA Laboratory[2]
Toulouse, France
[1]valerian.mange@hensoldt.fr
[2]valerian.mange@tesa.prd.fr

Jean-Yves TOURNERET
University of Toulouse, ENSEEIHT-IRIT[3]
TéSA Laboratory[2]
Toulouse, France
[3]jean-yves.tourneret@toulouse-inp.fr

François VINCENT
ISAE-Supaéro
Toulouse, France
francois.vincent@isae-supaero.fr

Laurent MIRAMBELL
Hensoldt Nexeya France
Toulouse, France
laurent.mirambell@hensoldt.fr

Fábio MANZONI VIEIRA
Hensoldt Nexeya France
Toulouse, France
fabio.manzoni@hensoldt.fr

***Abstract*—Multiple sensors, such as AIS and radar, are used to monitor nearby ships during maritime surveillance operations. The data from these sensors must be associated so as to accurately locate the targets and identify their behavior, while taking into account the presence of potential sensor biases. Several algorithms exist in the state-of-the-art to solve this association problem. However, few of them allow the sensor biases to be corrected. This paper adapts the coherent point drift method to the association of AIS and radar tracks while taking into account the radar uncertainty. The proposed adaptation is based on an expectation-maximization algorithm that jointly estimates the bias of the radar sensor with respect to the AIS sensor (in polar coordinates), the radar and AIS uncertainties and solves the association problem. The performance of this algorithm is evaluated using AIS and radar tracks obtained from numerous scenarios yielding promising results.**

***Index Terms*—AIS, Radar, Track, Association, Coherent Point Drift, Bias, Uncertainty**

I. Introduction

Context. Many operations are conducted by sea and are susceptible to disruption or can be sources of illegal activity. Thus, maritime surveillance missions have to be carried out to ensure the security of national interests. However, due to the large number of vessels, the analysis of vessel behavior cannot be done manually and must be automated [1], especially if multiple sensors are combined. Different sensors exist to track the ship movement, the common combination being AIS (Automatic Identification System) and radar. The motivation for such a combination is that AIS, while being precise and containing a large amount of information, is not mandatory on all vessels and can be illegally turned off or falsified. This paper addresses the problem of simultaneous association and re-calibration of radar and AIS data, which remains an important research topic.

Research supported by the French Army Ministry of Defence - Agence de l'innovation de défense.

State-of-the-art. Several association algorithms have been proposed in the literature [2], [3], [4], [5] among which Coherent Point Drift (CPD) [6] is a very efficient strategy. CPD is based on an Expectation-Maximization (EM) algorithm [7] that iteratively estimates an additive bias between two sets of points and the association probabilities of the observations.

Objectives, contributions and organization. This paper investigates a new EM algorithm for AIS and radar data association inspired by CPD. Instead of working in Cartesian coordinates, we propose to consider polar coordinates, allowing a possible angle mismatch between the AIS and radar sensors to be considered. Under these assumptions, closed-form expressions for the angular bias between AIS and radar and the uncertainties of the two sensors are obtained. This leads to a fast AIS and radar association algorithm, which is interesting for maritime surveillance applications. The EM algorithm requires a sufficiently accurate initialization of its parameters. This paper studies a new initialization maximizing a likelihood corresponding to assigning each radar observation to all the AIS data with the same probabilities.

The paper is organized as follows: Section II introduces the sensors that are considered in this work for maritime surveillance. Section III recalls the principles of the CPD algorithm and explains how to adapt this algorithm to the association of AIS and radar data. Section IV is devoted to the study of convergence of the CPD algorithm. Experiments illustrating the results of the proposed algorithm are given in Section V. Conclusions are reported in Section VI.

II. AIS and Radar for Maritime Surveillance

This paper investigates an algorithm allowing AIS and radar data to be associated with the aim of using the complementary properties of the two sensors (accuracy of the AIS system and reliability of the radar data). This section briefly reminds some properties of AIS and radar sensors before presenting the coordinate system considered in this work.

A. AIS

Though initially used to prevent collisions between ships, AIS has become a widespread system for monitoring vessels, with data constantly available everywhere in the world[1]. AIS messages provide a large amount of information, such as the position in Cartesian coordinates with an error less than 10 meters, speed and heading (both relative and absolute) of vessels, as well as vessel identifiers, dimensions, point of departure and arrival. The frequency of AIS varies from 2 seconds to 6 minutes, depending on the speed and maneuver [8]. Only ships above a certain size are required to emit AIS whereas it is only recommended for smaller vessels. Moreover, AIS messages can be non-available or subjected to spoofing. Thus, AIS information has to be complemented by another sensor, such as the radar presented in the next section.

B. Radar

This work assumes that radar data are available in addition to AIS tracks, providing the range and angle of the vessels contained in a common region of interest for both sensors. Uncertainties are also introduced for the radar data, modeling the fact that the further away a vessel is from the radar, the less accurate the tracks associated with this target. In polar coordinates, the uncertainty matrix for the radar is expressed as $\mathbf{\Sigma} = \text{diag}(\sigma_r^2, \sigma_\theta^2)$ where σ_r^2 and σ_θ^2 are the variances of the radar range and azimut, and "diag" transforms these variances into a diagonal matrix. The values of σ_r^2 and σ_θ^2 should reflect the fact that radars are more accurate in range than in azimut [9]. Furthermore, since the zero for the azimut has to be calibrated, a small bias exists along this dimension. While it is common to estimate associations and biases one after the other [4], this paper focuses on introducing a statistical model allowing these quantities to be estimated jointly.

C. Coordinate system

Before introducing the proposed statistical model, an appropriate representation of the positions has to be chosen. This work only considers ship positions but other features, such as speed and heading, could be added in the model to improve the association performance. Though tracks are often provided in Cartesian coordinates, both AIS and radar data are considered in this work in polar coordinates to build a simple Gaussian Mixture Model (GMM) [9]. In a given time window, we consider a set of AIS points $\mathbf{X}^{\text{AIS}} = [\mathbf{x}_1^a, \ldots, \mathbf{x}_K^a]$ and a set of radar points $\mathbf{X}^{\text{radar}} = [\mathbf{x}_1^r, \ldots, \mathbf{x}_N^r]$, where $\mathbf{x}_i^a$ and $\mathbf{x}_i^r$ contain the ith range r_i^a and azimuth θ_i^a of the AIS track, and the ith range r_i^r and azimut θ_i^r of the radar track, K is the number of AIS data and N is the number of radar data. The motivations for this choice will be detailed in Section III-A. Note that the AIS and radar data correspond to the same region of interest, which corresponds to the field of view of the radar.

[1] www.marinetraffic.com/en/ais/home/

III. AIS / Radar association

This section recalls the theory behind the CPD algorithm used to estimate the association probabilities and the unknown bias in Sections III-A1 and III-A2. The CPD theory is then extended to AIS / radar association with unknown uncertainty in Section III-A2. When sets of points need to be associated, the CPD algorithm uses one of these sets to define the means of a GMM and defines the covariance matrices of this GMM using some knowledge about the relationships between the different sets of points. The unknown parameters of the resulting GMM (including the associations probabilities) are then estimated using the EM algorithm.

A. Proposed CPD Model and Parameter Estimation

1) Statistical model: The two sets of points $\mathbf{X}^{\text{AIS}}$ and $\mathbf{X}^{\text{radar}}$ are assumed to contain the different positions (in polar coordinates) of the vessels observed by the AIS and radar at a fixed time t. As explained in Section II, the radar data are observed with a small angular bias denoted as θ that has to be estimated. Since AIS data are generally more precise than the radar, the following model is used for AIS/Radar data association:

$$p(\mathbf{x}_n^r) = p(\mathbf{x}_n^r|\mathbf{X}^{\text{AIS}}) = (1-w)\sum_{k=1}^{K} \pi_k p(\mathbf{x}_n^r|\mathbf{x}_k^a) + \frac{w}{\pi R^2}, \quad (1)$$

with $n = 1, ..., N$, $\pi_k = \frac{1}{K}$ the prior probability of $\mathbf{x_k^a}$ and

$$p(\mathbf{x}_n^r|\mathbf{x}_k^a) = \mathcal{N}(\mathbf{x}_n^r|[0,\theta]^T + \mathbf{x}_k^a, \mathbf{\Sigma}), \quad (2)$$

where θ is the unknown angular bias of the radar, $\mathcal{N}(\mathbf{x}|\mu, \mathbf{\Sigma})$ denotes the Gaussian distribution of mean μ with covariance $\mathbf{\Sigma}$ and $p(\mathbf{x}_n^r|K+1) = \frac{1}{\pi R^2}$ is the probability density function (pdf) of a uniform distribution on the region of interest, which is introduced to cope with all the outliers, as in [3]. Note that the weight w can be chosen by the user to reflect the quantity of outliers contained in the observed radar data. In the application targeted by this paper, the numbers of observed AIS and radar data at the instant t are known.

2) EM algorithm: After initializing the different parameters to be estimated (see Sections IV-C and V-B for details), the EM algorithm iterates between the following E and M steps:

- Expectation step: At a given iteration $l \geq 1$, the E step calculates the posterior probabilities $\gamma^l(z_{nk})$ with $z_{nk} = 1$ if data n belongs to the class k, 0 otherwise:

$$\gamma^l(z_{nk}) = \frac{\pi_k p^{l-1}(\mathbf{x}_n^r|\mathbf{x}_k^a)}{p^{l-1}(\mathbf{x}_n^r)}, \quad (3)$$

  where $p^{l-1}(\mathbf{x}_n^r|\mathbf{x}_k^a)$ and $p^{l-1}(\mathbf{x}_n^r)$ are obtained from (1) and (2) and the estimated parameters at iteration $l-1$.
- Maximization step: Once the posterior probabilities $\gamma^l(z_{nk})$ have been calculated, the model parameters are re-estimated during the M step. This is accomplished by maximizing the following Q function [7] with respect to the unknown parameters:

$$Q = -\sum_{n=1}^{N}\sum_{k=1}^{K+1} \gamma(z_{nk}^l) \ln\left[\pi_k p^l(\mathbf{x}_n^r|\mathbf{x}_k^a)\right]. \quad (4)$$

Straightforward computations allow the Q function to be expressed as:

$$Q = -\sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk}^{l})\left[\ln\frac{1}{2\pi K\sigma_r\sigma_\theta} - \frac{(r_n^r - r_k^a)^2}{2\sigma_r^2} - \frac{(\theta_n^r - \theta_k^a - \theta)^2}{2\sigma_\theta^2}\right]. \quad (5)$$

The minimum of (5) with respect to the bias has a closed form expression, since

$$\frac{\partial Q}{\partial \theta} = 0 \Rightarrow \hat{\theta} = \frac{\sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk}^{l})(\theta_n^r - \theta_k^a)}{\sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk}^{l})}. \quad (6)$$

Similarly, setting to zero the partial derivatives of Q with respect to the uncertainties σ_r^2 and σ_θ^2 leads to:

$$\frac{\partial Q}{\partial \sigma_r^2} = 0 \Rightarrow \hat{\sigma}_r^2 = \frac{\sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk}^{l})(r_n^r - r_k^a)^2}{\sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk}^{l})}, \quad (7)$$

$$\frac{\partial Q}{\partial \sigma_\theta^2} = 0 \Rightarrow \hat{\sigma}_\theta^2 = \frac{\sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk}^{l})(\theta_n^r - \theta_k^a - \hat{\theta}^l)^2}{\sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk}^{l})}. \quad (8)$$

The advantage of working in polar coordinates is clear since this system of coordinates leads to closed-form expressions for the unknown model parameters. Consequently, each iteration of the EM algorithm can be performed very fast, which is important for maritime surveillance.

B. Track-to-Track Association

The fuzzy associations given by the posterior probabilities calculated in the E step of the EM algorithm need to be converted into associations between AIS and radar data. The Hungarian algorithm [10] is applied to the association matrix $\mathbf{\Gamma} = [\gamma(z_{nk})]_{1\leq n\leq N, 1\leq k\leq K}$ obtained from the last iteration of EM, as in [3]. Any significant radar outlier, i.e., with $\gamma(z_{n(K+1)}) \geq \gamma(z_{nk}), \forall k = 1, ..., K$ is removed from the association by the Hungarian algorithm.

IV. Algorithm Convergence

Despite being a powerful tool, the EM algorithm requires a good initialization. This section first studies the sensitivity of the proposed EM algorithm to its initialization. An initialization avoiding many convergence issues is then proposed.

A. Simulation scenario

This section considers two datasets referred to as "Baltic" and "California". The Baltic Sea set (see Fig. 1) contains vessel positions that are clearly separated while the California set (Fig. 2) contains clusters of vessels, which makes it more difficult to handle. For both of these sets, only the AIS data is available and was acquired from IEEEDataPort[2] for the Baltic Sea set and from MarineCadastre[3] for the California set. The radar dataset was simulated from the AIS trajectories according to the statistical model considered in this paper.

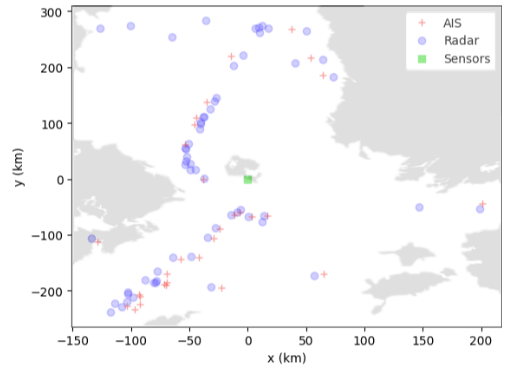

Fig. 1. Surveillance zone using the Baltic Sea set (AIS observations: red crosses, radar observations: blue circles, sensors: green square).

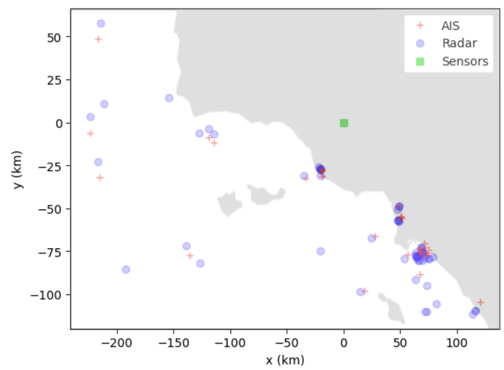

Fig. 2. Surveillance zone using the California set (AIS observations: red crosses, radar observations: blue circles, sensors: green square).

For each AIS track defined in polar coordinates, a bias θ has been added to the AIS azimut coordinate and zero-mean Gaussian noises have been added to the biased azimut and range trajectories. Finally, the ship positions are taken at a time instant to apply the CPD algorithm. In both experiments, the parameters are set to $w = 0.1$, $\sigma_r = 25.51$ meters, $\sigma_\theta = 0.1531°$ (this corresponds to 50 meters and 0.3° with a confidence of 95%) with 100 scenarios generated. Each scenario uses 60 ship trajectories: 30 trajectories with both AIS and radar positions and 30 trajectories with only radar positions (i.e., $K = 30$ and $N = 60$). The number of iterations for the EM algorithm is set to 10.

B. Sensitivity to the initialization step

To study the sensitivity of the EM algorithm to the initialization step for AIS and radar association, a series of experiments has been performed. This section studies a specific scenario with an actual angular radar bias $\theta = 4°$. For each experiment, the initial estimate of the bias is chosen randomly between $-45°$ and $45°$ and $N_{\text{Mc}} = 200$ Monte Carlo runs are performed. Fig. 3 shows the boxplots of the EM estimates for the Baltic and California datasets. The EM algorithm does not always converge towards the global maximum of the likelihood for a random initialization. Initializing the bias with $\theta = 0°$ seems to be a decent first guess, as the radar should have a small bias. However, this initialization does not always lead to

[2] ieee-dataport.org/open-access/vessel-tracking-ais-vessel-metadata-and-dirway-datasets

[3] marinecadastre.gov/accessais/

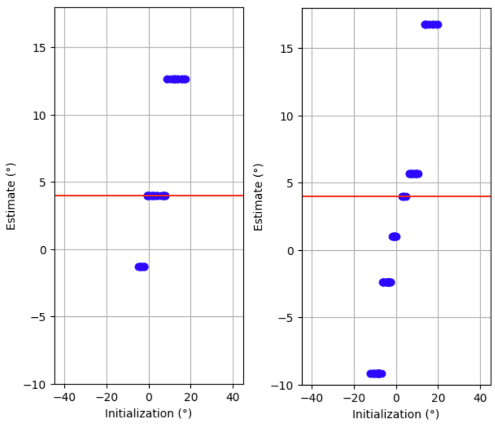

Fig. 3. Bias estimates obtained using different initializations with the Baltic Sea (left) and California (right) sets.

a global maximum. To overcome this issue and decrease the execution time (i.e., reducing the number of iterations required for the EM algorithm), a more clever initialization needs to be found, as proposed in the next section.

C. Proposed initialization

As explained in [11], when using the EM algorithm to determine the means of each component of a GMM, the global maximum of the likelihood is reached if two criteria are met:

- the target Gaussian means are all clearly separated,
- the initialization of the Gaussian centers is close enough to the target values.

As explained in Section III, the AIS positions in the proposed statistical model define the Gaussian centers of the GMM. The proposed initialization consists of finding the value of θ (by grid search) that maximizes the cost function

$$c(\theta) = \sum_{n=1}^{N} \ln \left(\sum_{k=1}^{K} \mathcal{N}(\mathbf{x}_n^r | [0, \theta]^T + \mathbf{x}_k^a, \mathbf{\Sigma}) \right), \quad (9)$$

which consists of assigning each radar observation to all the AIS data with the same probabilities.

V. Simulation results

This section evaluates the performance of the proposed method in two different situations corresponding to known and unknown uncertainties (for known uncertainties, the variance estimates in (7) and (8) are removed from the EM algorithm and these parameters are replaced by their known values).

A. Known uncertainty

The estimation performance obtained for known uncertainties (known values of σ_r^2 and σ_θ^2) is displayed in Figs. 4 and 5 using box plots of the estimates. First, one can observe that the EM algorithm improves the bias estimation compared to the initial estimate calculated using (9). Note that the final estimate is close to the actual value of θ with an error close to $0.01°$ in the vast majority of simulations for the California set and with an error close to $0.001°$ for the Baltic Sea set. The Hungarian

TABLE I
Associations for the Baltic Sea (B) and California (C) sets with known (K) and unknown (U) uncertainties

	BK	BU	CK	CU
Averaged Correct Associations (%)	100	99.97	97.73	94.80
Standard deviation (%)	0	0.33	2.25	3.69

algorithm applied to the matrix $\mathbf{\Gamma}$ provides the associations given in Table I. These associations are perfect (no error) for the Baltic Sea set and correspond to a probability of correct association above 90% for the California set.

B. Unknown uncertainty

In the case of unknown uncertainties, the parameters σ_r^2 and σ_θ^2 are estimated jointly with the radar bias θ. The initial values of these variances are generated equally likely in the intervals $[10\text{m}, 50\text{m}]$ and $[0.01°, 0.26°]$. The bias θ is initialized randomly between $-10°$ and $10°$. These intervals for σ_r^2 and σ_θ^2 correspond to values that can be expected from real radars[4] and AIS data with a confidence of 95%.

Compared to the previous results, a minor loss in performance can be observed in Figs. 4 and 5 (right plots) for the bias estimation and in Table I for the associations. This loss of performance was expected since adding two unknowns to the problem also increases its complexity. Figs. 6 and 7 (left plots) show the uncertainty estimations obtained with EM that are compared to the true value. Though the results seem acceptable for the Baltic Sea set, the uncertainties for the California set have higher variance (even without outliers) and are overestimated. This overestimation is probably due to association errors resulting from incorrect values of the posterior probability associations $\gamma(z_{nk})$. Indeed, according to (7) and (8), non-binary association probabilities tend to give some weight to the neighbors of the actual vessels instead of considering a single association. The fuzzy associations for the Baltic dataset are close to be hard decisions, which allows better uncertainty estimates to be obtained. In order to confirm this assessment, the uncertainties have been re-estimated using the associations provided by the Hungarian algorithm. Figs. 6 and 7 (right plots) show the gain in performance obtained when using the posterior probabilities resulting from the Hungarian algorithm for uncertainty estimation.

VI. Conclusion

This work investigates a new statistical model for track-to-track association from AIS and radar data. One specificity of the proposed model is to allow a bias correction for the radar data and AIS / radar uncertainty estimation within the association algorithm. Simulations conducted on realistic datasets showed promising results, with a good precision for the radar bias estimates and a probability of correct association greater than 95% in most cases, even with a dense traffic.

[4] www.radartutorial.eu/01.basics/Radars%20Accuracy.en.html

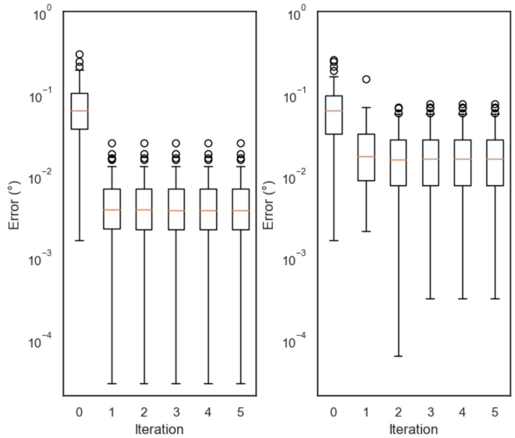

Fig. 4. Angle estimation errors after multiple iterations for the Baltic Sea set with (left) and without (right) uncertainties.

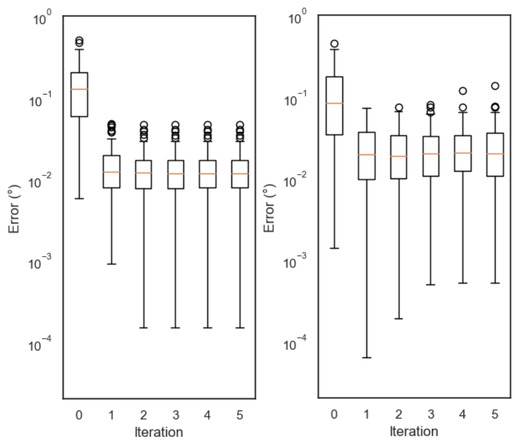

Fig. 5. Angle estimation errors after multiple iterations for the California set with (left) and without (right) uncertainties.

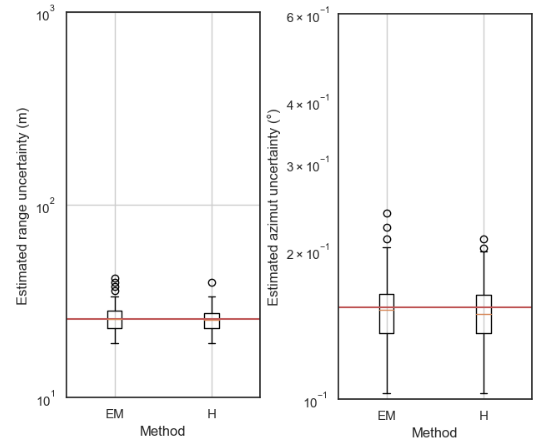

Fig. 6. Uncertainty estimates for the Baltic Sea set obtained using EM (EM) and Hungarian (H) associations with the target value in red.

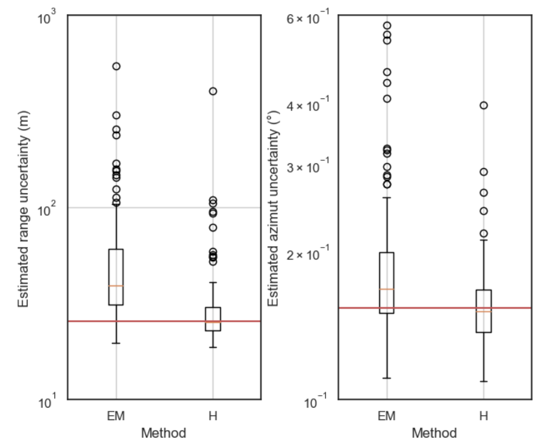

Fig. 7. Uncertainty estimates for the California set obtained using EM (EM) and Hungarian (H) associations with the target value in red.

However, the presence of outliers in the reference dataset has not been handled, even if these are unlikely to affect the results when computing the posterior probabilities. Future work will be devoted to including electronic support measure (ESM) into the analysis. ESM is a passive sensor measuring angles from incoming signals in a specific bandwidth, usually corresponding to radar usage. ESM have shown interesting properties for maritime surveillance, such as identifying friend and foe or the emitter used by neighboring ships during operations [12].

References

[1] J. Venskus, P. Treigys, J. Bernataviciene, G. Tamulevicius, and V. Medvedev, “Real-time maritime traffic anomaly detection based on sensors and history data embedding,” *Sensors*, pp. 1–10, Aug. 2019.

[2] Y. Bar-Shalom and H. Chen, “Multisensor track-to-track association for tracks with dependent errors,” in *Proc. Int. Conf. on Decision and Control (CDC)*, Paradise Island, Bahamas, Dec. 2004, pp. 2674–2679.

[3] H. Zhu, M. Wang, K.-V. Yuen, and H. Lueng, “Track-to-track association by coherent point drift,” *IEEE Signal Processing Letters*, vol. 24, no. 5, pp. 643–647, March 2017.

[4] J. Wang, Y. Zeng, S. Wei, Z. Wei, Q. Wu, and Y. Savaria, “Multi-sensor track-to-track association and spatial registration algorithm under incomplete measurements,” *IEEE Transactions on Signal Processing*, vol. 69, pp. 3337–3350, May 2021.

[5] S. Fortunati, F. Gini, A. Farina, A. Graziano, and Giompapa S. Greco, M. S., “On the application of the expectation-maximisation algorithm to the relative sensor registration problem,” *IET Radar, Sonar and Navigation*, vol. 7, no. 2, pp. 191–203, Feb. 2013.

[6] A. Myronenko and X. Song, “Point set registration: Coherent point drift,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2262–2275, March 2010.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer New York, 2006.

[8] R. Sturgis, V. Emiya, B. Couëtoux, and P. Garreau, “Beyond geofencing: Behavior detection using AIS,” *Ocean Engineering*, vol. 293, pp. 116630, Jan. 2024.

[9] D. Cormack, I. Schlangen, J. R. Hopgood, and D. E. Clark, “Joint registration and fusion of an infrared camera and scanning radar in a maritime context,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 2, pp. 1357–1369, April 2020.

[10] J. Munkres, “Algorithms for the assignment and transportation problems,” *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, no. 1, March 1957.

[11] R. Zhao, Y. Li, and Y. Sun, “Statistical convergence of the EM algorithm on Gaussian mixture models,” *Electronic Journal of Statistics*, vol. 14, no. 1, pp. 632–660, Jan. 2020.

[12] T. Pietkiewicz, A. Kawalec, and B. Wajszczyk, “Analysis of fusion primary radar, secondary surveillance radar (IFF) and ESM sensor attribute information under Dezert-Smarandache theory,” in *Proc. Int. Radar Symposium (IRS)*, Prague, Czech Republic, June 2017, pp. 1–10.