Enhancing maritime situational awareness through multimodal fusion: insights from a real-world experiment

Enhancing Maritime Situational Awareness through Multimodal Fusion: Insights from a Real-World Experiment

1st Kilian Wohlleben
Center for Digital Safety & Security
Austrian Institute of Technology GmbH
Vienna, Austria
kilian.wohlleben@ait.ac.at

2nd Michael Hubner
Center for Digital Safety & Security
Austrian Institute of Technology GmbH
Vienna, Austria
michael.hubner@ait.ac.at

3rd Thanet Markchom
Computational Vision Group
University of Reading
UK
thanet.markchom@reading.ac.uk

4th Jonathan Boyle
Computational Vision Group
University of Reading
UK
j.n.boyle@reading.ac.uk

5th James Ferryman
Computational Vision Group
University of Reading
UK
j.m.ferryman@reading.ac.uk

6th Stephan Veigl
Center for Digital Safety & Security
Austrian Institute of Technology GmbH
Vienna, Austria
stephan.veigl@ait.ac.at

7th Andreas Opitz
Center for Digital Safety & Security
Austrian Institute of Technology GmbH
Vienna, Austria
andreas.opitz@ait.ac.at

8th Anastasios Gkamaris
SKYLD
Cyprus
a.gkamaris@skyld.com.cy

9th Romeo Bratskas
SKYLD
Cyprus
rb@skyld.com.cy

***Abstract*—Effective maritime border surveillance is crucial. Challenges we face include irregular migration, smuggling, oil spills and the need for rapid search and rescue. Various sensing technologies, including AIS, SAR, optical and infrared sensors, as well as UAV-mounted sensors, clearly enhance maritime awareness. However, integrating their diverse outputs remains complex. Feature-level multi-modal sensor fusion is a well-known methodology for robust detection and behavior analysis. However, most research relies on simulations or isolated sensors, which limits practical insights.**

This study presents a controlled real-world experiment combining synchronized data from coastal ground sensors and UAV-mounted visual and infrared sensors. The recorded dataset enables the evaluation of feature-level fusion in authentic conditions. We enhance existing fusion frameworks with additional modules and assess them using operational metrics. This study contributes to our understanding of the efficacy of multi-modal fusion in complex maritime environments, while also highlighting the significant challenges involved in transitioning from simulations to controlled real-world sensor data.

***Index Terms*—Maritime Surveillance, Sensor Data Fusion, Situational Awareness, Fusion Applications, Real-World dataset**

I. Introduction

In recent years, maritime border surveillance has gained significant geopolitical relevance, driven by growing concerns about irregular immigration, smuggling activities, environmental incidents such as oil spills, and the ongoing requirement for efficient search and rescue operations. Given that maritime zones are vast, dynamic and often hostile environments, both on the surface and underwater, ensuring persistent and effective surveillance remains a major challenge. A detailed review of the challenges and future directions in maritime security is reported in [1].

A wide range of sensing technologies has been developed and successfully implemented in this domain [2]. These include traditional coastal and ship-based systems such as Automatic Identification Systems (AIS), Synthetic Aperture Radar (SAR) [3]–[5], electro optical (EO) sensors [6], and infrared (IR) sensors in both the short-wave and long-wave spectrum. In addition, satellite systems [7] offer high-altitude perspectives for large-area monitoring and allow change detection over time. Meanwhile, unmanned aerial vehicle (UAV) platforms are being used more and more for specific missions.

However, this technological diversity presents its own challenges. These heterogeneous data sources operate under different environmental conditions and generate vast amounts of information that are often difficult to interpret in isolation. As

This project received funding from the European Union's Horizon Europe research and innovation programme under grant agreement no. 101073985.

Funded by
the European Union

a result, multi-modal sensor fusion has emerged as a necessary paradigm for combining complementary data streams and enhancing the detection of maritime entities [8], [9] (e.g., ships, small vessels, and individuals) in complex operational settings, as well as supporting behavior analysis. The scientific community has made notable advances in fusion methodologies [10]–[14] particularly at feature level - demonstrating that generalized approaches of multi-modal data fusion can significantly improve overall surveillance performance.

Despite these advances, a significant limitation persists: many fusion strategies are evaluated on simulated data [8], [9] or address the challenges at sensor level [11] (e.g.; pixel level fusion) [15]. Although such studies are an essential step in the development process, they often fail to capture the complexities and uncertainties of real-world sensor interactions. Suitable, synchronized, multi-sensor datasets for feature-level fusion are scarce, often due to the classified nature of the data. This limits reproducibility and comparability in this critical area of research.

In this paper, we present an evaluation study explicitly designed to address this gap. We describe a real-world experiment in which we recorded and analyzed synchronized data from coastal ground sensors and UAV-mounted sensors for feature-level fusion in maritime border surveillance scenarios. Building upon the fusion modules presented in [14], we evaluate their performance, adaptability, and real-world relevance across a set of operational metrics. However, our main contribution lies in presenting key findings on the transition from simulations to controlled real-world sensor data.

The structure of this paper is as follows: Section II describes the experimental setup and data acquisition methodology. Section III elaborates on the fusion architecture, and the evaluation metrics used in this study. Section IV presents the experimental results, while Section V reflects on the operational implications and key findings as well as the future work. Finally, Section VI concludes the paper by summarizing the main contributions.

II. Experimental Setup

A. Sensors

The following is a list of sensor platforms used to generate geolocated detections:

Coastal Ground Sensors: The coastal ground sensor system comprises two infrared cameras mounted on poles about six meters high. The short wave infrared (SWIR) camera [16] is equipped with fixed optics that provide a 15° field of view (FOV) for detecting small vessels up to 800 meters away and people up to 200 meters away. The long wave infrared (LWIR) camera [17] is equipped with configurable optics that allow the user to adjust the zoom for an optimal field of view based on operational requirements.

UAV-EO: The aerial platform uses an advanced octocopter to provide real-time target detection. It is equipped with an RGB camera and an onboard NVIDIA Jetson Orin 64, which runs the detection model. The primary sensor is an electro-optical camera that operates in the visible spectrum (approximately 400–700 nm) and has a 30x optical zoom. This provides stable imagery during flight. Its optical parameters include a 4K sensor resolution (3840 x 2160), a 1/1.32-inch CMOS sensor, a native frame rate of 30 fps, a focal length range of 6.72–201.6 mm, a field of view ranging from 63.7° (wide) to 2.3° (telephoto), and an aperture range of f/1.6–f/4.7.

UAV-IR: This platform uses a DJI Mavic 3T Enterprise [18] equipped with a custom Android application to stream real-time telemetry and thermal video data to a ground server. The UAV's thermal camera has a diagonal field of view (DFOV) of 61°, operating within the infrared wavelength range of 8–14 μm. Thermal video is recorded at a resolution of 640×512 pixels at 30 frames per second.

GNSS Trackers: For the collection of the ground truth (i.e timestamped locations of vessels and persons) GNSS trackers by [19] (eTrex® SE) and Septentrio [20] were used. Each tracker records GNSS locations in WGS84 coordinates, along with a timestamp for each second. The Garmin trackers served as the primary devices, while the Septentrio trackers were used as backup devices in case of incomplete data.

B. Scenario descriptions

Six scenarios were selected from a live demonstration that was recorded over two days. These scenarios depict interactions between people, vessels, and vehicles. Visual representations of each scenario are provided. Fig. 1.

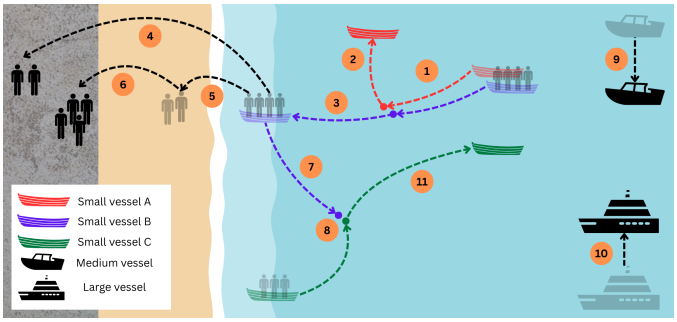
Scenarios 1-3 : This scenario features three small vessels (A, B, and C), one medium vessel, one large vessel, and multiple people (see Fig. 1a).

The scenario begins with small vessels A and B heading toward the shore (1). Vessel B is carrying a group of people, while vessel A leads the way. When both vessels are approximately 20 meters from the shore, vessel A turns around and departs (2), leaving vessel B to continue alone (3).

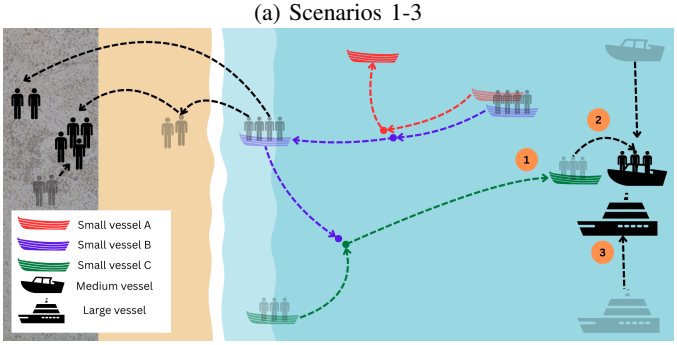
At approximately 10 meters from the shore, four individuals disembark from vessel B and wade through the water to the beach. Two of them continue running inland (4), while the other two remain on the beach (5). After the first pair reaches the inland area, the remaining two follow and meet a separate group of people (not the first pair) waiting there (6). After the people disembark, vessel B turns and moves away from the shore in a different direction from its initial path (7). Small vessel C, also carrying people, then approaches and meets with vessel B (8).

Meanwhile, the medium and large vessels, previously stationary in the outer waters, begin moving. The medium vessel travels parallel to the shore (9), while the large vessel moves toward it, also parallel to the shoreline (10). Small vessel C then leaves small vessel B, heading rapidly toward the medium vessel (11).

Scenarios 4-6: These extend the sequence from scenario 3: beginning when small vessel C starts heading toward the medium vessel (see Fig. 1b) until it reaches the medium vessel (1). Small vessel C reaches the medium vessel before the large vessel, and small vessel C's passengers board the medium vessel (2). Shortly after, the large vessel enters the scene and



(a) Scenarios 1-3



(b) Scenarios 4-6

Fig. 1: Visual representations of each scenario in the experiments.

positions itself near the medium vessel as the scenario ends (3).

C. Sensor execution

This section describes the movement patterns of the sensor platforms for each scenario (as described in section II-B. The ground sensor platforms are limited to orientation changes, whereas both UAV platforms have full movement within the operational area.

Scenarios 1-3: GNSS trackers were placed on small vessel A, small vessel B, small vessel C, the medium vessel, and one person on board small vessel B who later disembarked and moved inland on foot. The SWIR and LWIR cameras followed small vessel B until it met small vessel C, then shifted focus to small vessel C, and finally to the medium vessel. The UAV-EO focused on small vessel B the entire time. The UAV-IR initially followed small vessel B, then tracked the disembarking individuals as they moved toward the shore. When the first pair ran inland, the UAV-IR followed them until they reached their destination. The UAV-IR then returned to capture and track the second pair running inland, until they met with the separate group of people. Afterwards, the UAV-IR left the inland area, moved back toward the sea, and focused on small vessel C.

Scenarios 4-6: GNSS trackers were deployed on the small vessel A, small vessel B, small vessel C, and large vessel. Sensor movements were the same as in Scenario 3 up to the point where small vessel C approached the medium vessel. From that moment onward, all sensors shifted their focus to the medium vessel and followed its movement until the end of the scenario.

D. Collected Evaluation Data

Each scenario, each deployed sensor platform captured the scenario play-out from a different angle and produced detections based on its field of view. The number of detections produced by each platform for each scenario is shown in Table I. Additionally, the GNSS trackers recorded ground truth data for each scenario. Table I also includes the number of data points recorded by the GNSS trackers.

TABLE I: Evaluation data

Scenario	#Detections				#GNSS	Time
	SWIR	LWIR	UAV-EO	UAV-IR		
1	35172	18303	1052	242	2248	12:41
2	32565	17345	514	127	2900	15:41
3	17118	14543	66	85	1241	7:43
4	19023	14811	97	86	937	6:33
5	19315	13397	129	105	1132	8:43
6	16322	18580	104	110	1075	8:16

III. Methodology

In this study we select a representative architecture which is described in [14]. We reiterate on the Hierarchical Fusion Graph (HFG) and explain how we apply and evaluate this architecture to the maritime domain in a controlled real-world environment. To improve readability and completeness we summarize the concept of a HFG and present the nodes that where used in this experimental setup.

A. Fusion architecture

The methodology described in [14] introduces the Hierarchical Fusion Graph (HFG), which is designed to model and construct multi-modal sensor data fusion systems. A fusion graph $\mathcal{G}$ is formulated as a directed acyclic graph (DAG), in which nodes and edges represent the data flow (edges) and processing models (nodes) of a fusion system. The graph comprises three distinct types of nodes: namely, sensor nodes S, fusion nodes F, and decision nodes D. The relation between different nodes follow specific rules. For example all sensor nodes $s_i \in S$ serve as the sources of data and have no incoming edges. Fusion nodes $f_i \in F$ aggregate information from one or more predecessors whereas both the sensor and fusion nodes possess at least one successor. Subsequently, decision nodes $d_i \in D$ serve to finalize the data processing. These nodes are considered terminal, implying that they must be preceded by at least one other node and do not output to any subsequent nodes. For a more detailed description we refer to [14]. In this study we focus on the application of this methodology in our experiment.

1) Sensor Nodes: The sensor information is modeled using Probabilistic Occupancy Maps (POM), which incorporate sensor observations via a grid containing the likelihood of an event occurring at a specific region of the map. The parameters of a POM are defined by the tuple:

$$(\tau_i, p_i) \quad \forall s_i \in S$$

The parameter p_i denotes the sensor prior imposed on observations, whereas the parameter τ_i models the temporal decay of information in the map. In the present experiment, the following detection algorithms were utilized as inputs for the sensor nodes:

Coastal Ground Sensor Model: A YOLOX [21] object detection algorithm was employed for detection on both coastal ground sensors. The direction of the bounding boxes of the detected persons and small vessels was estimated using the extrinsic camera parameters for the purpose of geo-localization. The geo-location of the model, as determined by the sensor observations, is represented by a cone (see Fig. 2-dark blue) with a maximum range of 800 meters. The opening angle was estimated from the width of the bounding box. The confidence of the sensor observation was estimated based on the softmax score..

UAV-EO Model: The detection algorithm utilizes a Single Shot Multibox Detector [22] (SSD) model with MobileNetV2 [23] backbone for real-time detection of people, vehicles and vessels. Per-object confidence is produced by the SSD classifier head as class probabilities after non-maximum suppression at a threshold of 0.5. The system excludes detections estimated to be more than 1 km from the platform based on monocular geo-projection using UAV GNSS/IMU pose, altitude, camera intrinsic, and gimbal angles resulting in circular geometries modelling the observation's geo-location (see Fig. 2).

UAV-IR Model: A retrained YOLOv11 [24] model detects objects (person, vessel, and vehicle) from the thermal video feed, and their geo-locations are estimated by using a geometric algorithm based on UAV telemetry and camera parameters. The platform outputs detection results where each detection includes the object class, confidence score, and estimated geo-location based on the intrinsic GNSS/IMU pose, altitude, camera intrinsic, and gimbal angles (see Fig. 2).

2) Fusion Nodes: In the context of the fusion nodes, the Bayesian fusion model is employed, as outlined in [14]. Furthermore, we introduce two new nodes. Let $\mathcal{M} = \{\mathbf{M}_1, \mathbf{M}_2, \dots \mathbf{M}_K\}$ be a set of POMs. A cell m_{ij} of $\mathbf{M}_k$ with probability p of being occupied at time t is represented in log-odds form by $l_t^k(m_{ij}) = \log \frac{p}{1-p}$. Then we define a Bayesian fusion by

$$B_t^{1:K}(m_{ij}) = \sum_{k=1}^{K} l_t^k(m_{ij}), \tag{1}$$

the LogicalAND fusion by:

$$\wedge_t^{1:K}(m_{ij}) = \min\{l_t^k(m_{ij}) : k \in 1..K\}, \tag{2}$$

and consequently the LogicalOR (not used in this study) by

$$\vee_t^{1:K}(m_{ij}) = \max\{l_t^k(m_{ij}) : k \in 1..K\}. \tag{3}$$

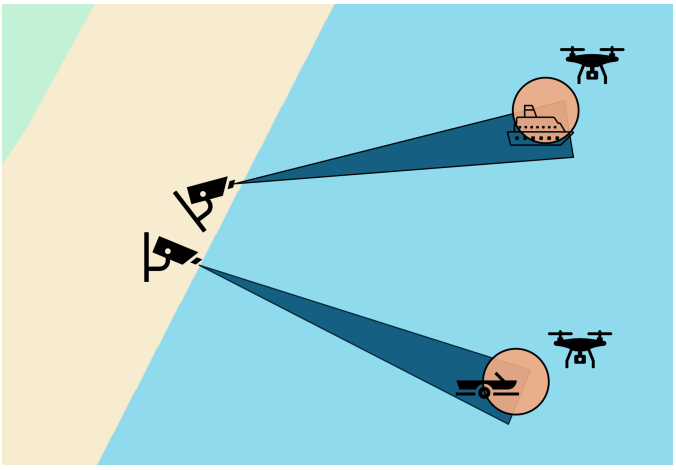


Fig. 2: Illustration of geo-located sensor observations of ships. UAV EO/IR - orange cirlces. Cameras SWIR/LWIR - Dark Blue cones

3) Decision Nodes: The decision node utilizes the thresholding method as outlined in [14], which involves the transformation of the log-odds employed in the fusion process back into probabilities. Subsequently, a threshold, denoted by θ, is applied, and image processing is employed to identify connected components. This process generates geo-localized fused observations that represent the output of a fusion graph $\mathcal{G}$.

B. Evaluation

In this section, the evaluation methodology is introduced, including a description of the data and the metrics that are used for the evaluation. The evaluation is limited to the ship detection component of the scenarios, as not all individuals were equipped with trackers, which would have resulted in erroneous classification as false positives. Given that our use case encompasses the identification and tracking of vessels, we assess the F1-Score, the False Positive Rate (FPR), and the distance between sensor/fusion detections and the ground truth. The calculation of the F1-Score and the FPR for geo-spatial data is described in [14]. For the distance metric we compute the distance of the centroid of the detections. The distance metric is computed by determining the centroid of the detections. For each ground truth, the minimal distance is assigned as its distance value. Then, the average is computed across all ground truths.

C. Selection of Fusion Graphs

One of the main tasks in constructing fusion systems is deciding how to combine different sensors to exploit their complementary data. In [14] the authors argue that this can be done by selecting different fusion graphs. Based on the scenarios and platform movement described in II-B we decided to investigate two fusion graphs $\mathcal{G}_1$ and $\mathcal{G}_2$. The graph $\mathcal{G}_1$ represents a fusion of sensor observations by mobile platforms (i.e. UAV-EO and UAV-IR), while $\mathcal{G}_2$ additionally incorporates

TABLE II: Parameters for the simple fusion graph.

	Standard P.	Optimized P.
UAV-EO (τ, p)	(1, 0.5)	(1.7424, 0.1654)
UAV-IR (τ, p)	(1, 0.25)	(0.1705, 0.0036)
θ	0.8	0.6847
Resolution (m)	5	5

costal ground sensors (i.e. SWIR and LWIR). The graphs are presented in Fig. 3 and Fig. 4 respectively, using the notation described in III-A2. These graphs were selected to assess the performance of mobile sensors versus simple fusion first, and then to compare their performance with the addition of coastal ground sensors. The first graph, or $\mathcal{G}_1$, uses a simple fusion approach that combines two similar sensors and performs low-pass filtering through thresholding. This allows us to compare the fusion system with the sensors alone to establish a baseline for further evaluations. The Logical AND fusion model combines the mobile sensors with the coastal-based cameras in the graph $\mathcal{G}_1$, requiring a detection from both sensor types to trigger an alarm. This combines the best aspects of both sensor types: the drones' precise localization capabilities and the cameras' reliability.

IV. Results

In this section we show the results of the evaluated metrics of the selected fusion graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ applied on the scenarios described in II-B. The evaluation is two-folded. First, we assess the sensors on the mobile platforms as a baseline. We only use the UAV data as a baseline. This enables us to compare the fusion graphs to the system not utilizing any fusion component. Next, we evaluate the graph $\mathcal{G}_1$ using standard (baseline) parameterization and an optimized version. The optimal parameters were found by running a Bayesian optimization as described in [25] with a target function $F_1 - \frac{Distance}{100}$, which was arbitrarily chosen based on experience and the use case of tracking in mind. Finally, we evaluate the full graph $\mathcal{G}_2$. In Tab. IV we see the final results of the evaluation of the selected metrics for the different fusion graphs by scenario and averaged. For the sake of completeness and reproducibility in Tab. II we show the final parameters of the fusion graph's nodes. Comparing the metrics of sensor observations from UAVs only to the fusion system represented by the graph $\mathcal{G}_1$ we see a reduction in the average over all scenarios of the FPR by three percentage points while the the F1-Score remained marginally unchanged, however, at the cost of the localization accuracy by roughly 2m. The results of the fusion system represented by $\mathcal{G}_1$ with optimized node parameter seem to only improve the F1-score compared to the baseline parametrization. Interestingly, by incorporating also the coastal ground sensor data represented by the graph $\mathcal{G}_2$ we do see a major reduction in the FPR at the cost of a slight reduction in the F1-score and distance. An optimization of the camera parameters was out of scope of this paper, as each step of the optimization requires a full run of our fusion system and, since there were many more camera than UAV detections (as seen in Tab. I), it was not feasible to run such an optimization.

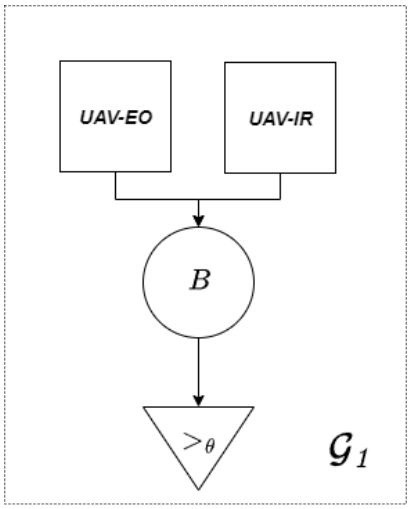


Fig. 3: Composition of Fusion Graph $\mathcal{G}_1$

TABLE III: Results of the different scenarios with different fusion configurations.

	UAV Data	$\mathcal{G}_1$	$\mathcal{G}_1$ Optimized	$\mathcal{G}_2$
FPR	0.1247	0.0991	0.1174	**0.0415**
Scenario 1	0.0618	0.0342	0.0403	0.0124
Scenario 2	0.1638	0.1491	0.1774	0.0379
Scenario 3	0.1125	0.0962	0.1043	0.0119
Scenario 4	0.1639	0.1367	0.1283	0.0788
Scenario 5	0.1623	0.0965	0.1389	0.0879
Scenario 6	0.0841	0.0820	0.1151	0.0202
F1-Score	0.3814	0.3850	**0.4062**	0.3644
Scenario 1	0.6341	0.6273	0.6638	0.5481
Scenario 2	0.2650	0.2762	0.2784	0.2611
Scenario 3	0.2892	0.2953	0.3038	0.2434
Scenario 4	0.3921	0.4039	0.4319	0.3929
Scenario 5	0.3429	0.3457	0.3670	0.3579
Scenario 6	0.3650	0.3616	0.3924	0.3831
Distance (m)	**6.3692**	8.6460	8.7942	6.9874
Scenario 1	4.8917	7.0977	6.8001	6.7353
Scenario 2	7.8228	10.7120	10.2642	8.9884
Scenario 3	6.2837	8.2908	8.5270	6.6501
Scenario 4	6.3519	9.1729	9.2113	6.2012
Scenario 5	6.7334	8.5465	9.1140	6.3347
Scenario 6	6.1315	8.0559	8.8487	7.0149

TABLE IV: Parameters of fusion for graph $\mathcal{G}_2$.

Parameter	Value
UAV-EO (τ, p)	(1, 0.5)
UAV-IR (τ, p)	(1, 0.5)
SWIR (τ, p)	(0.2, 0.3)
LWIR (τ, p)	(0.2, 0.3)
θ	0.8
Resolution (m)	5

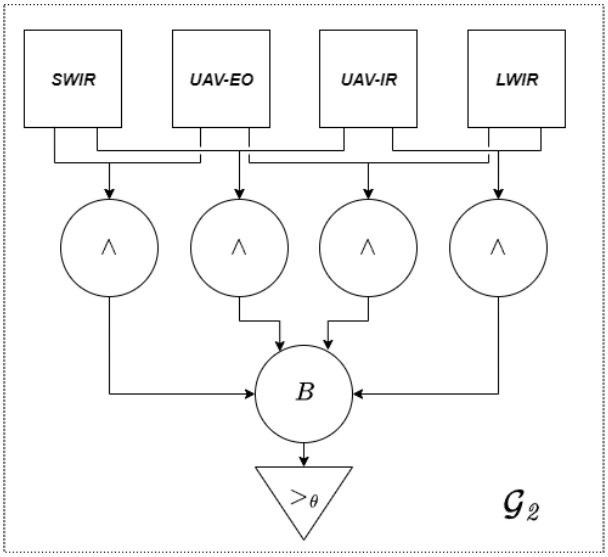


Fig. 4: Composition of Fusion Graph $\mathcal{G}_2$

V. Discussion

Developing an experimental setup for scientific research in these domains presents many challenges, both technical and organizational. In this study, we also reflect on this perspective because we believe it provides valuable insights into developing such an experiment. However, our focus remains on the technical and scientific findings we gained by conducting and evaluating our fusion models within this experiment.

A. The Operational Perspective

In section II, we described the experimental setup. Planning is pivotal for conducting such an experiment successfully. The focus was on generating representative data for multimodal fusion systems in the context of maritime surveillance. The use cases in this context are complex and require many resources, such as small boats, vessels, actors, and, of course, a location for the experiment. This comes with a financial burden. Despite the existence of funded research projects, generating a complete data set for scientific benchmarking of fusion systems is nearly impossible due to the classification, sensitivity, and multiple stakeholder involvement. The silver lining is that it is still possible to report the findings, which help the community share lessons learned. This is how we hope to contribute with this manuscript.

Data Collection Table I summarizes the final data set. However, this does not represent all of the collected data, but rather the data after curation. In real-world experiments, you are exposed to environmental influences (bad weather, wind) and technical issues (malfunctioning experimental equipment, power outages, etc.) that may compromise the experiment temporarily. Ultimately, only 58:17 minutes of the recorded data from a two-day experiment was suitable for evaluating the fusion system of our experiment.

Synchronicity One of the key characteristics of this dataset was its synchronous nature. In previous trials, we found that synchronizing all the data sources was very difficult when the sensors were connected via different networks and the computers were set to different time zones. Another common issue is the slight offset between UTC and GPS time, which many people are unaware of and is difficult to detect in the data. It is impossible to distinguish a slight time error from an error in space with the naked eye. To ensure temporal alignment, we used an NTP server to which all the sensor platforms and sensors were connected. Furthermore, we synchronized the GNSS trackers, meaning the sensor detections and ground truth data are in the correct temporal order. This was necessary for the temporal domain of our evaluation.

B. The Technical Perspective

Geo-Localization: Before discussing the results of the evaluation, we would like to reiterate the importance of geolocating sensor information. Since the application is geospatial, modeling the location of sensor observations is crucial. This is especially important because the spatial domain is part of the evaluation metric, as seen in other experiments [12]–[14], [25]. From a domain perspective, covering large areas is important, as is the distance to the nearest object of interest. However, as shown in Tab. III, we observe that the scaling of the distance metric is rather small (less than 10 m). For large areas, this is quite accurate, considering an operator needs to locate a ship's position. However, the coastal-based cameras weren't positioned high enough to provide accurate distance estimates for their detections.

Results in General: Even though the metrics don't look super impressive, the UAV detections themselves were already very good. This is because, for our evaluation, even a small divergence leads to a false positive classification. For example, if a wind gust affects the UAV and it locates the ship 15 meters away from its actual location, this leads to an incorrect detection. A visual inspection of the data revealed no false positives; the camera only detected ships when they were present. The coastal-based sensors performed similarly, except they sometimes detected ships on the horizon that weren't part of our scenario. Our goal was to reduce the FPR while maintaining the F1-score and distance. Due to the high quality of the data, any fusion system can only slightly influence the metrics, and a trade-off is always involved. Fusion of UAVs compared to UAVs alone showed an improvement in FPR and F1-score, confirming the benefits of a fusion system as argued in [14]. The optimization of the graph parameter does not seem impressive. Although the F1-score increased, the FPR decreased compared to the baseline, due to the selected target metric. The full graph is the best option for sequences defined in scenarios. From an operational point of view, the small differences in distance and recall are not that relevant. Since the vessels were detected most of the time, operators still have a good overview of the situation. The decrease in FPR is much more significant because false alarms are usually costly and, in our case, would lead to an overestimation of the number of ships in the area. Using two different sensor modalities to confirm the presence of a ship was beneficial. Thus, we

can confirm that using fusion systems to handle and exploit sensor data is necessary for better performance and situational awareness.

Limitations: There are still limitations that need to be considered. First, fusion systems rely on the underlying data. This means that in case none of the sensors are capable of any detection this capability will not magically appear after the fusion system. However, as we have seen in Tab. IV fusion systems are capable of exploiting the positive attributes of sensors and thus contribute to the performance enhancing measured by different metrics. Therefore, adding more data sources (i.e., sensors) to the sensor network increases the potential of fusion systems. Second, constructing and parameterizing (i.e., configuring) a fusion system remains very complex. In this study, the fusion graph was selected empirically based on knowledge of the scenarios being planned. This is considered domain know-how and is only available through experts in the field. This knowledge cannot be assumed a priori and is difficult to obtain in certain domains. This is why domain-agnostic optimization techniques are necessary. In our future work, we plan to investigate such methodologies by extending the optimization process described in [25] to incorporate the selection of appropriate nodes (i.e., sensor, fusion, and decision).

Working with real-world data: Several real-world operational limitations affected the evaluation. For instance, ships outside the defined experimental area were detected, contributing to the observed false positives. Full coverage could not be achieved due to operational procedure constraints, resulting in a lower absolute F1-score. These conditions underscore the importance of a robust fusion system capable of handling false positives, which are often triggered by technical artifacts and environmental phenomena in real-world sensor deployments.

Ultimately, we decided against evaluating the coastal ground sensor alone, as the cameras also picked up ships on the horizon that weren't part of the experiment. This means that the results would not be representative. This decision is important, as it demonstrates how the experimental setup is influenced by changing environmental conditions in a real-world scenario. As there were vastly more camera detections, as seen in Tab. III, evaluating the full fusion graph $\mathcal{G}_2$ is very costly. Because of these reasons we decided not to optimize the whole fusion graph as we did with $\mathcal{G}_1$. Quite frankly, we also did not anticipate significant benefits from the optimization results of $\mathcal{G}_1$.

C. Key Takeaways & Future Work

Clearly, state-of-the-art methodologies perform reliably on synthetic datasets. However, this experiment demonstrates that real-world environments are complex, characterized by noisy, incomplete, and often unsynchronized sensor data. Unexpected detections, such as vessels not part of the experiment, revealed limitations in spatial coverage and presented challenges in interpreting, curating, and synchronizing the data. Furthermore, constraints specific to sensors, such as directional outputs from optical cameras or degraded GNSS signals affecting UAV-based geolocation, highlight technical limitations in observation modeling. This experiment definitively demonstrates the challenges of working with controlled real-world sensor data. Multi-modal sensor fusion is the clear solution to these challenges. By leveraging the strengths of different sensors, it improves detection metrics. However, achieving consistently high performance remains difficult due to imprecise and incomplete data. Performance is also heavily influenced by metric selection, which varies with application needs.

Applying a fusion graph to the maritime domain has shown promise in improving localization robustness, indicating that fused data could support more sophisticated techniques. The goal of using this dataset is to make multi-target tracking easier to implement in future research. It is also hoped that valuable real-world insights will be obtained when these methodologies are applied. However, identifying an optimal fusion graph is a challenging task that usually requires domain expertise. Subsequent efforts will focus on developing techniques to automatically select or adapt graph structures based on performance metrics or by encoding domain knowledge directly into the fusion process. Furthermore, the experiment corroborates the importance of multimodality. The sensor suite is planned to expand to include AIS, SAR, and high-altitude sources, such as satellite images. This expansion will mitigate individual sensor deficiencies and enhance the system's overall resilience.

VI. Conclusion

This study offered valuable insights from a real-world experiment in maritime surveillance. The study emphasized the challenges and advantages of transitioning fusion applications from simulated environments to controlled real-world scenarios. Planning such an endeavor is complex. This is especially true when considering a multi-technology provider and multi-authority experiment. One of the most time-consuming tasks is securing resources, such as the availability of ships, vessels, and actors. Our primary contribution is demonstrating how applying a hierarchical fusion graph-based approach to a sensor network composed of coastal ground sensors and UAV-mounted sensors can enhance system performance and situational awareness. The two main ship movement scenarios evaluated in this study demonstrate the critical role of fusion systems in effectively integrating heterogeneous sensor data. Our findings emphasize that applying fusion methodologies to real-world maritime data presents significant challenges. These challenges are technical, operational, and related to deployment and coordination. These insights provide a valuable foundation for advancing practical, fusion-based maritime surveillance systems.

References

[1] Z. Bhuiyan, R. Hasan, I. Gocer, S. Ahmad, and Z. Bhuiyan, "AI in Maritime Security: Applications, Challenges, Future Directions, and Key Data Sources," July 2025. https://www.preprints.org/manuscript/202507.0113/v1 - visited on 2025-10-27.

[2] M. Guerriero, P. Willett, S. Coraluppi, and C. Carthel, "Radar/AIS data fusion and SAR tasking for Maritime Surveillance," in *2008 11th International Conference on Information Fusion*, pp. 1–5, June 2008.

[3] M. Rodger and R. Guida, "Classification-Aided SAR and AIS Data Fusion for Space-Based Maritime Surveillance," *Remote Sensing*, vol. 13, p. 104, Jan. 2021. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[4] S. D. Priyadharshini and K. Vadivazhagan, "Enhanced Vessel Detection in Maritime Surveillance using Multi-Modal Data Integration and Deep Learning," in *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 1090–1099, Oct. 2024. ISSN: 2768-0673.

[5] D. Nikolic, N. Stojkovic, Z. Popovic, N. Tosic, N. Lekic, Z. Stankovic, and N. Doncov, "Maritime Over the Horizon Sensor Integration: HF-SWR Data Fusion Algorithm," *Remote Sensing*, vol. 11, p. 852, Jan. 2019. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

[6] R. W. Liu, Y. Guo, J. Nie, Q. Hu, Z. Xiong, H. Yu, and M. Guizani, "Intelligent Edge-Enabled Efficient Multi-Source Data Fusion for Autonomous Surface Vehicles in Maritime Internet of Things," *IEEE Transactions on Green Communications and Networking*, vol. 6, pp. 1574–1587, Sept. 2022.

[7] G. Soldi, D. Gaglione, N. Forti, L. M. Millefiori, P. Braca, S. Carniel, A. D. Simone, A. Iodice, D. Riccio, F. C. Daffinà, D. Quattrociocchi, G. Bottini, P. Willett, and A. Farina, "Space-Based Global Maritime Surveillance. Part II: Artificial Intelligence and Data Fusion Techniques," *IEEE Aerospace and Electronic Systems Magazine*, vol. 36, pp. 30–42, Sept. 2021.

[8] A. Gad and M. Farooq, "Data fusion architecture for Maritime Surveillance," in *Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002. (IEEE Cat.No.02EX5997)*, vol. 1, pp. 448–455 vol.1, July 2002.

[9] H. Wu, J. Xian, X. Mei, Y. Zhang, J. Wang, J. Cao, and P. Mohapatra, "Efficient target detection in maritime search and rescue wireless sensor network using data fusion," *Computer Communications*, vol. 136, pp. 53–62, Feb. 2019.

[10] W. Huang, H. Feng, H. Xu, X. Liu, J. He, L. Gan, X. Wang, and S. Wang, "Surface Vessels Detection and Tracking Method and Datasets with Multi-Source Data Fusion in Real-World Complex Scenarios," *Sensors*, vol. 25, p. 2179, Jan. 2025. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

[11] Y. Fischer and A. Bauer, "Object-oriented sensor data fusion for wide maritime surveillance," in *2010 International WaterSide Security Conference*, pp. 1–6, Nov. 2010. ISSN: 2166-1804.

[12] M. Hubner, K. Wohlleben, M. Litzenberger, S. Veigl, A. Opitz, S. Grebien, and M.-T. Dvorak, "A Bayesian Approach - Data fusion for robust detection of vandalism and trespassing related events in the context of railway security," in *2024 27th International Conference on Information Fusion (FUSION)*, pp. 1–7, July 2024.

[13] M. Hubner, K. Wohlleben, M. Litzenberger, S. Veigl, A. Opitz, S. Grebien, F. Graf, A. Haderer, S. Rechbauer, and S. Poltschak, "Robust Detection of Critical Events in the Context of Railway Security Based on Multimodal Sensor Data Fusion," *Sensors*, vol. 24, p. 4118, Jan. 2024. Publisher: Multidisciplinary Digital Publishing Institute.

[14] M. Hubner, J. Nausner, and K. Wohlleben, "Advanced Sensor Fusion for Railway Security - A Hierarchical Graph-Based Approach," in *2024 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pp. 1–7, Nov. 2024. ISSN: 2473-7666.

[15] J. Zhan, J. Li, L. Wu, J. Sun, and H. Yin, "VIOS-Net: A Multi-Task Fusion System for Maritime Surveillance Through Visible and Infrared Imaging," *Journal of Marine Science and Engineering*, vol. 13, p. 913, May 2025. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

[16] "Atlas SWIR 1.3MP IP67 Model (IMX990) - LUCID Vision Labs," Sept. 2021. https://thinklucid.com/product/atlas-swir-1-3mp-model-imx990/ - visited on 2025-08-14.

[17] "AXIS Q8752-E Bispectral PTZ Camera - Product support | Axis Communications." https://www.axis.com/products/axis-q8752-e/ - visited on 2025-08-14.

[18] "Technische Daten - DJI Mavic 3 Enterprise Serie - DJI Enterprise." https://enterprise.dji.com/de/mavic-3-enterprise/ - visited on 2025-08-14.

[19] Garmin and G. L. o. i. subsidiaries, "Garmin eTrex® SE | Handheld Hiking GPS." https://www.garmin.com/en-US/p/835742/ - visited on 2025-08-14.

[20] "GNSS receiver manufacturer for precise positioning." https://www.septentrio.com/en/ - visited on 2025-08-14.

[21] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021," Aug. 2021. arXiv:2107.08430 [cs].

[22] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," vol. 9905, pp. 21–37, 2016. arXiv:1512.02325 [cs].

[23] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 2017. arXiv:1704.04861 [cs].

[24] G. Jocher and J. Qiu, "Ultralytics yolo11," 2024. https://github.com/ultralytics/ultralytics.

[25] K. Wohlleben, F. Siems, J. Nausner, and M. Hubner, "Bayesian optimization for parameter selection in fusion systems," in *2025 28th International Conference on Information Fusion (FUSION)*, pp. 1–7, 2025.