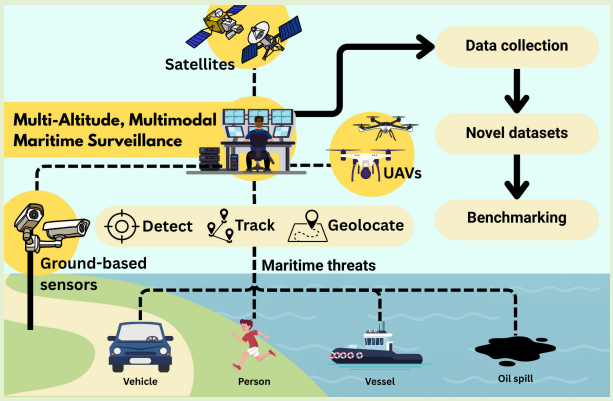
Multi-altitude, multimodal maritime surveillance system

Markchom, T. ORCID: https://orcid.org/0000-0002-2685-0738, Kourounioti, O., Marturini, M., Bratskas, R., Wohlleben, K., Boyle, J. ORCID: https://orcid.org/0000-0002-5785-8046, Chen, L., Voskopoulos, G., Kontopoulos, C., Veigl, S., Opitz, A., Gkamaris, A., Papachristos, D., Lunic, D., Ferryman, J., Kriechbaum-Zabini, A. and Leventakis, G. (2026) Multi-altitude, multimodal maritime surveillance system. IEEE Sensors Journal, 26 (5). 7101 -7119. ISSN 1558-1748 doi: 10.1109/JSEN.2025.3649549 Available at https://centaur.reading.ac.uk/127545/

Multialtitude, Multimodal Maritime Surveillance System

Thanet Markchom, Olympia Kourounioti, Matteo Marturini, Romaios Bratskas, Kilian Wohlleben, Jonathan Boyle, Lulu Chen, George Voskopoulos, Christos Kontopoulos, Stephan Veigl, Andreas Opitz, Anastasios Gkamaris, Dimitris Papachristos, Dumitru Lunic, James Ferryman, *Member, IEEE*, Andreas Kriechbaum-Zabini, and George Leventakis

***Abstract*—Maritime surveillance plays a vital role in protecting coastal and maritime environments. However, traditional maritime surveillance systems that rely on single-altitude, single-modality sensors suffer from limited coverage and sensitivity to weather conditions. To address these limitations, this article presents a comprehensive maritime surveillance system that integrates multialtitude, multimodal sensor platforms, including ground-based sensors, low-altitude uncrewed aerial vehicles (UAVs), and satellites, for maritime threat detection. Each platform is equipped with dedicated modules for object detection, tracking, and geolocation, leveraging its unique sensing capabilities to contribute to a coordinated surveillance system. Moreover, a novel multialtitude, multimodal maritime surveillance (MAMMS) dataset is introduced. This dataset includes data from these sensor types, enabling rigorous benchmarking across varying operational conditions. The experimental results indicate that the system achieved an average mAP of 50.5% across all sensors in object detection, surpassing state-of-the-art models in most cases. For object tracking, the system achieved an average ID F1-Score (IDF1) of 0.263 and a higher order tracking accuracy (HOTA) of 0.297, comparable to state-of-the-art methods, while exhibiting substantially fewer average ID switches (IDSWs) (75.46) compared to the strongest baseline (301.46). For geolocation approximation, the system achieved an error of less than 11 m in certain scenarios. A case study was also conducted to assess the sensor platforms when integrated into a multisensor fusion system. The case study showed that complementary information from different platforms can help reduce false alarms and improve object geolocation accuracy. The dataset is available at https://zenodo.org/records/17979190**



***Index Terms*—Geolocation approximation, ground-based sensors, maritime surveillance, multialtitude sensors, multimodal sensors, multiple-object tracking, object detection, satellite remote sensing, uncrewed aerial vehicle (UAV).**

I. Introduction

MARITIME surveillance is a crucial aspect of ensuring the safety, security, and environmental protection of coastal and maritime areas. With increasing maritime activities and the growing concern over illegal activities, there is a pressing need for more effective and comprehensive surveillance systems capable of monitoring vast maritime regions and detecting potential threats in real time [1].

Traditional maritime surveillance approaches primarily rely on coastal radar stations [2], automatic identification system (AIS) transponders [3], and optical cameras mounted on ground-based fixed platforms or ships [4], [5], [6], [7], [8], [9]. While useful for situational awareness, these systems have limitations: AIS requires cooperative tracking and fails with unauthorized vessels, radar has limited open-sea coverage, and optical cameras are affected by poor weather and low visibility [10].

Many studies have explored the deployment of visual sensors with diverse modalities across a range of platforms to extend the capabilities of traditional maritime surveillance systems [11]. These include multimodal ground-based cameras [11], [12], vessel-mounted sensors [13], low-altitude aerial vehicles [14], [15], and satellites [16], [17]. Despite

Received 8 December 2025; accepted 14 December 2025. Date of publication 13 January 2026; date of current version 2 March 2026. This work was supported by the European Union's Horizon Europe Research and Innovation Project EURMARS: An Advanced Platform to Improve European Multi Authority Border Security Efficiency and Cooperation (https://eurmars-project.eu/), under Agreement 101073985. The associate editor coordinating the review of this article and approving it for publication was Dr. Muhammad Ali Jamshed. *(Corresponding author: Thanet Markchom.)*

Please see the Acknowledgment section of this article for the author affiliations.

Digital Object Identifier 10.1109/JSEN.2025.3649549

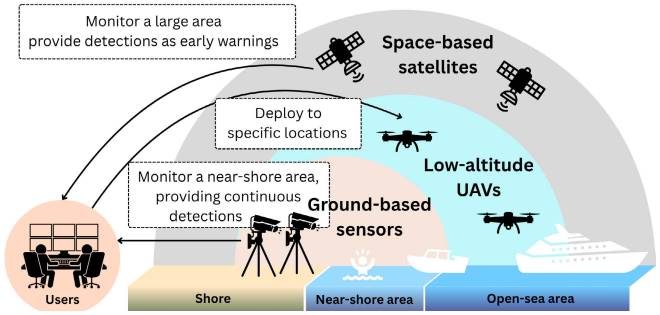


Fig. 1. Overview of the proposed multialtitude, multimodal surveillance system that integrates multimodal ground-based sensors, low-altitude UAVs, and space-based satellite sensors.

these advances, most existing visual multimodal maritime surveillance systems remain limited by their focus on a single operational altitude. For example, systems may deploy low-altitude aerial vehicles in isolation or rely solely on satellite imagery, resulting in fragmented situational awareness.

This work addresses these limitations by proposing a comprehensive maritime surveillance system that integrates multialtitude, multimodal sensor platforms: ground-based multimodal systems, low-altitude uncrewed aerial vehicles (UAVs), and space-based satellite sensors. Each platform is equipped with modules for object detection, tracking, and geolocation estimation, tailored to its sensing characteristics and operational role within the surveillance architecture. Fig. 1 illustrates an overview of the proposed system and how sensor platforms at different altitudes cooperate. By combining the high-resolution and persistence of ground-based sensors, the flexibility and mobility of UAVs, and the broad coverage of satellites, the proposed system forms a coordinated network capable of robustly detecting and localizing maritime threats with enhanced situational awareness.

Moreover, to facilitate the evaluation of the proposed system, this article also introduces a novel multialtitude multimodal maritime surveillance (MAMMS) dataset that combines data from the aforementioned sensor platforms. This dataset enables systematic benchmarking of key perception tasks, including object detection, long-term tracking, and geolocation estimation, across multiple altitudes and sensing modalities. This benchmarking offers valuable insights into the strengths and limitations of each sensor type and highlights their contributions to the overall surveillance framework. The key contributions of this article can be summarized as follows.

1) Propose a multialtitude, multimodal maritime surveillance system integrating ground-based, low-altitude UAV, and satellite sensors, each equipped with object detection, tracking, and geolocation estimation modules.
2) Introduce a novel dataset combining data from sensor types with different altitudes and modalities.
3) Benchmark the performance of each sensor platform on object detection, long-term tracking, and geolocation approximation tasks using the proposed dataset.

II. Related Work

Visual sensors with various modalities have been explored in maritime surveillance systems. In [12], a pipeline for multiobject detection and tracking was proposed, combining optical and thermal video from ground-based cameras to support persistent coastal monitoring. A vessel detection model introduced in [13] fused RGB and infrared (IR) images from on-vessel cameras. Similarly, Zhan et al. [11] presented a method that enhances ship type recognition using co-aligned visible and IR images from a dual-mode ground-based sensor. Low-altitude platforms have also been considered. For instance, Jang et al. [14] proposed a detection framework that integrates high-resolution RGB with low-resolution hyperspectral images collected from utility aircraft. Satellite-based methods have also gained traction. The work in [16] combined synthetic aperture radar (SAR) and optical satellite imagery to improve vessel detection, while Priyadharshini and Vadivazhagan [17] fused SAR, optical, and IR data together with AIS information to increase detection robustness. Despite these advances, most existing studies focus on leveraging multiple sensors with different modalities at a single altitude. A significant gap remains in the development of maritime surveillance systems that effectively integrate multialtitude, multimodal sensor platforms.

In addition to developing multisensor maritime surveillance systems, evaluating their performance is equally important. To support this, many studies have introduced multimodal datasets for surveillance and threat detection. For instance, MarDCT [18] includes videos from fixed, moving, and pan–tilt–zoom (PTZ) cameras, while VAIS [19] provides synchronized RGB and long-wavelength IR (LWIR) ship images for cross-modality classification. IPATCH [4], [20] combines visible and thermal imagery with global positioning system (GPS), AIS, and radar data for detection and tracking. Singapore Maritime [6] combines on-shore and onboard RGB and near-IR (NIR) videos. Seagull [21] and XMCMT [12] include UAV-based multispectral imagery and dual-modality videos with tracking annotations. M^2SODAI [14] provides synchronized RGB and hyperspectral imagery (HSI) from aircraft. QXS-SAROPT [16] consists of satellite SAR and optical imagery. WaterScenes [22] integrates radar, monocular camera, GPS, and inertial measurement unit (IMU) sensors for detection and segmentation tasks. Collectively, these datasets span various modalities and platforms. However, most are limited to a single-altitude level, lacking the integration of ground-based, low-altitude, and space-based sensors. This highlights the need for multialtitude, multimodal datasets to enable more comprehensive evaluation of maritime surveillance systems.

III. Scopes and Problem Statement

The threats considered in this work are defined by a range of objects that may pose risks to maritime and coastal safety, security, and the environment. These include: 1) persons (either on deck, on land, or in the water), 2) vessels, 3) vehicles, and 4) oil spills (significant environmental threats).

To detect such threats, sensor platforms in maritime surveillance systems should: 1) detect objects and classify them as persons, vessels, vehicles, or oil spills; 2) track detected objects by assigning unique tracking IDs (if an object temporarily disappears due to occlusion or movement, it should be reassociated with the same ID when it reappears);

3) approximate the geolocations (latitude and longitude) of detected objects. This work focuses on developing multialtitude, multimodal sensor platforms to support these key tasks as part of an integrated maritime surveillance system.

IV. Multialtitude, Multimodal Maritime Surveillance System

This work explores four sensor platforms across different altitudes and modalities: 1) AIT Smart Sensor platform, which integrates visual RGB, thermal, SWIR, and UV sensors; 2) Thermal UAV platform, which integrates low-altitude thermal imaging sensors; 3) RGB UAV platform, which integrates visual RGB sensors operating at low altitude; and 4) Satellite-based systems, which include Sentinel-1 (SAR interferometry) and Sentinel-2 [only the visual bands (RGB)]. Each platform is equipped with dedicated models/methods for object detection, tracking, and geolocation approximation, tailored to its respective modality. Details of the individual platform are provided in this section.

A. AIT Smart Sensor Platform

For ground-based sensors, this article presents an advanced sensor platform designed for maritime and land-based surveillance called the **AIT Smart Sensor platform** [see Fig. 2(a)]. This platform is a mast-mounted, steerable system with AI-based detection and short-term tracking, and has been enhanced to improve detection and tracking performance. The system can identify and track individuals and vehicles up to 200 m on land and vessels up to 800 m at sea. The platform incorporates a ground-based visible RGB (**GS-RGB**), UV (**GS-UV**), thermal (**GS-Therm**), and SWIR (**GS-SWIR**) camera system mounted on a pan–tilt unit (PTU) for directional control, enabling comprehensive scanning of coastal and open-sea areas. Video streams are transmitted via a proprietary Gigabit Ethernet protocol to an algorithmic server, where AI-based detection and tracking modules process the data.

1) Object Detection: Two detection modules are developed for this platform: one for land detection and another for sea detection. Both are based on YOLO-X, modified and configured for the maritime environment. In particular, we conducted experiments on the collected data described in Section V and fine-tuned the hyperparameters of the pretrained YOLO-X model (YOLOX-l) to perform person, vessel, and vehicle detection in a maritime environment. The original classification head of YOLO-X is retained. Nonrelevant predictions are excluded, while those labeled as car, bus, or truck are remapped to the vehicle class, and those labeled as boat or vessel are remapped to the boat class. A REST interface is provided for configuring and controlling the module.

2) Object Tracking: The tracking algorithm follows the detect-and-track approach. Instead of using manually initialized bounding boxes, it relies on detection outputs from a backbone detector, YOLO-X, in our case. In the first frame, tracked objects are initialized based on the detector's output. In subsequent frames, the algorithm compares existing tracked objects with new detections by computing a distance matrix, where each element represents the Euclidean distance between a tracked object and a new detection. The optimal matching is then determined by solving a global assignment problem using the Hungarian algorithm. The hyperparameters that can be modified to adapt the algorithm's behavior are 1) the maximum number of consecutive frames in which an object may remain unobserved before it is removed from the track, and 2) the maximum distance between an existing tracked object and a new detection to be considered a valid association to the track. After empirical validation on selected training sequences (see Section V), these two hyperparameters were set, respectively, to 4 for the maximum number of consecutive frames and 0.6 for the distance.

3) Object Geolocation Approximation: The AIT Smart Sensor platform also includes an integrated global navigation satellite system (GNSS) receiver. It can be used to automatically determine the position and elevation of the platform. As the GNSS module contains several internal receiving antennas and a magnetometer, the direction of view of the sensors can also be determined accurately, in addition to their position.

Nevertheless, there is a fundamental problem in determining the exact position of an object with only one sensor position. The installation height of the AIT Smart Sensor platform is very low (2 and 3 m above the ground) compared to the distance of the objects (100–500 m). If a triangle consisting of the height of the mast, the angle of inclination of the sensors, and the plane section were to be used to determine the position, the flat angle would result in a very abrasive section. This means that the smallest deviation of the detection location in the image, or measurement errors in the sensor position and/or angle, can lead to large errors in the calculated geo-position. This, in turn, would lead to an unstable and "jumping" geo-localization. For this reason, the specification of an exact detection location is deliberately avoided, and it is instead represented as a sector/area (polygon).

Taking into account the detections in the image and the FOV of the sensors, a geo-localized triangular sector (polygon) can be calculated in which the detected object is located. The apex of the triangle corresponds to the camera's position. The apex angle, which defines the spread of the sector, is derived from the angular width of the detection in the image. This angular width is computed from the object's bounding-box width and the known horizontal field of view (HFOV) of the camera as follows:

$$\Delta\theta_{\text{box}} = \frac{w_{\text{box}}}{W} \times \text{HFOV} \tag{1}$$

where $\Delta\theta_{\text{box}}$ is an angular width of the detection, w_{box} is the bounding-box width, and W is the image width. The height of the triangle, measured from the camera to the base, is determined by a predefined depth according to the use case. The base of the triangle is formed by extending the two rays corresponding to the angular boundaries from the camera position until they intersect this depth, producing two vertices that define the polygon. In the maritime environment, the simplified assumption that all relevant objects are located on a single level (the water surface) is applied.

(a) (b) (c)

Fig. 2. Ground-based and low-altitude sensors (a) AIT Smart Sensor. (b) Thermal UAV. (c) RGB UAV.

B. Thermal UAV Platform

For a thermal UAV sensor (**UAV-Therm**), a commercially available DJI Mavic 3T [see Fig. 2(b)] is used with a custom Android application developed based on the DJI Mobile Software Development Kit (DJI Mobile SDK) version 5.[1] The application is installed on the remote controller (DJI RC Pro) and is designed to stream telemetry and imagery data from the UAV to a ground-based server in real time. To transmit telemetry data, the application uses socket communication for real-time transmission of essential flight and aircraft telemetry, including altitude, GPS coordinates, gimbal orientation, velocity, and battery level. For imagery data, the UAV provides live thermal video feeds at a resolution of 640 × 480, which are streamed through a real-time messaging protocol (RTMP) server to a processing machine.

1) Object Detection: YOLOv11 [23], the state-of-the-art YOLO model, is employed to develop an object detector for thermal UAV imagery. A pretrained YOLOv11 model (YOLO11s) is retrained using the data collected in this work (see Section V). The original 80-class COCO classification head in the model is retained, with our annotations mapped to compatible classes: person, vessel, and vehicle are converted to person, boat, and car. To retrain the model, the YOLO11s[2] pretrained weights were adopted. The retraining was performed for 50 epochs with a batch size of 16 and an input image size of 640 × 640. The learning rate was set to 0.01 with a weight decay of 0.0005. The loss function was weighted by 0.75 for the box loss and 0.5 for the classification loss. The dropout rate was set to 0. The default data augmentation strategy and hyperparameters[3] were adopted. As for inference, predictions are filtered to five relevant classes: person, boat, car, bus, and truck. Boats are converted back to vessels, while cars, buses, and trucks are considered vehicles.

2) Object Tracking: For the thermal UAV, BoT-SORT [24] is used as a tracker. This method offers improvements over previous SORT-based trackers by refining the Kalman Filter state and incorporating camera motion compensation through image registration. This makes it well-suited for UAV imagery, where the sensor is often in motion. In this setup, detections are provided by the proposed detection model (retrained YOLOv11) instead of the original YOLOX used in BoT-SORT. The hyperparameters were tuned using the training data for object tracking described in Section V. After tuning, the adopted configuration used a high-confidence threshold of 0.25 and a low-confidence threshold of 0.1. New tracks were initiated when the detection confidence reached 0.25, and existing tracks were retained for up to 30 frames in the absence of matching detections. The association similarity threshold was set to 0.8. Sparse optical flow was used for global motion compensation, and the proximity and appearance thresholds for identity matching were set to 0.5 and 0.8, respectively.

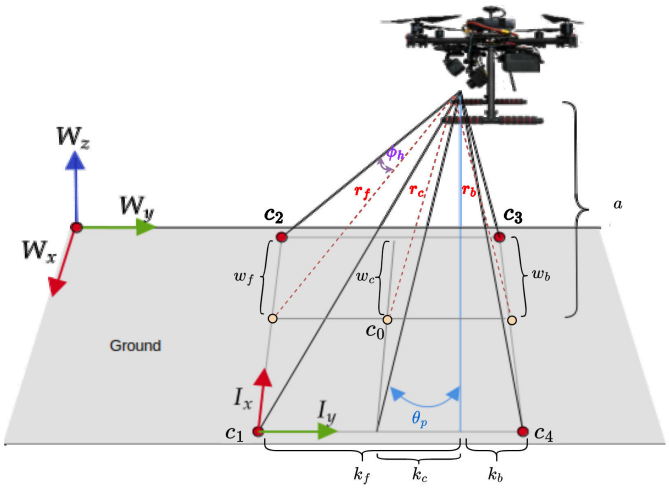


Fig. 3. Geolocation approximation algorithm for the thermal UAV and RGB UAV platforms.

3) Object Geolocation Approximation: To approximate an object's real-world coordinates, the ground projection of the UAV's field of view (FOV) is first computed. Subsequently, the geospatial coordinates of the four-corner points are determined. Finally, a homography transformation matrix is derived using these points as reference points to transform pixel coordinates into geospatial coordinates. Fig. 3 illustrates the projection geometry.

To estimate the ground projection of the UAV's FOV, the ground distances corresponding to the image center (k_c), front (k_f), and back (k_b) are first calculated. Given the UAV's altitude a, its camera pitch angle θ_p, and half of the vertical FOV angle ϕ_v, the distances are derived using the tangent

[1] https://github.com/dji-sdk/Mobile-SDK-Android-V5

[2] https://docs.ultralytics.com/models/yolo11/#supported-tasks-and-modes

[3] https://docs.ultralytics.com/modes/train/#augmentation-settings-and-hyperparameters

function as follows:

$$k_c = \frac{a}{\tan(|\theta_p|)}, k_f = \frac{a}{\tan(|\theta_p| + \phi_v)}, k_b = \frac{a}{\tan(|\theta_p| - \phi_v)}. \tag{2}$$

These distances represent the projections of the image center, front, and back onto the ground plane. The corresponding hypotenuse distances, which account for the camera's altitude, are then computed as

$$r_c = \sqrt{a^2 + k_c^2}, \quad r_f = \sqrt{a^2 + k_f^2}, \quad r_b = \sqrt{a^2 + k_b^2}. \tag{3}$$

Using these values, the half-widths of the image frame in ground coordinates at the center, front, and back are determined by

$$w_c = r_c \cdot \tan(\phi_h), w_f = \frac{r_f \cdot \tan(\phi_h)}{\text{ccf}}, w_b = \frac{r_b \cdot \tan(\phi_h)}{\text{ccf}} \tag{4}$$

respectively, where ϕ_h is half of the horizontal FOV angle, and ccf is a corner correction factor due to sensor crop and fisheye correction.

With these values, the center and four image corners in the global coordinate (latitude–longitude) system, incorporating the drone's position (d_{lat}, d_{lon}) and yaw angle θ_y, are then computed

$$\begin{aligned}
c_{0,lat} &= d_{lat} + w_c \cos(\theta_y) + k_c \sin(\theta_y) \\
c_{3,lat} &= d_{lat} + w_b \cos(\theta_y) + k_b \sin(\theta_y) \\
c_{4,lat} &= d_{lat} - w_b \cos(\theta_y) + k_b \sin(\theta_y) \\
c_{2,lat} &= d_{lat} + w_f \cos(\theta_y) + k_f \sin(\theta_y) \\
c_{1,lat} &= d_{lat} - w_f \cos(\theta_y) + k_f \sin(\theta_y).
\end{aligned} \tag{5}$$

Similarly, the longitude coordinates are computed as

$$\begin{aligned}
c_{0,lat} &= d_{lon} + k_c \cos(\theta_y) - (w_c) \sin(\theta_y) \\
c_{3,lon} &= d_{lon} + k_b \cos(\theta_y) - (w_b) \sin(\theta_y) \\
c_{4,lon} &= d_{lon} + k_b \cos(\theta_y) + (w_b) \sin(\theta_y) \\
c_{2,lon} &= d_{lon} + k_f \cos(\theta_y) - (w_f) \sin(\theta_y) \\
c_{1,lon} &= d_{lon} + k_f \cos(\theta_y) + (w_f) \sin(\theta_y).
\end{aligned} \tag{6}$$

To map any pixel coordinates from the image plane to geospatial coordinates, a homography matrix **H** is computed using the RANSAC-based robust method.[4] This transformation is derived by relating key reference points in the image (four-corner points) to their corresponding locations in the ground plane. After obtaining the transformation matrix, the bottom center point of each bounding box is taken as the representative point. Its pixel coordinate is then projected through **H** to obtain the approximated geolocation.

However, it is common that the UAV may operate at a low altitude and capture images from an oblique angle. This reduces projection accuracy, as the upper image region may include sky instead of ground. Consequently, the top corners may no longer correspond to valid ground points, leading to projection errors. To mitigate this, a horizon-aware ground projection correction is applied. First, the minimum pitch angle $\theta_{p,\min}$ at which the horizon aligns with the top edge of the image is determined. If the actual pitch $\theta_p > \theta_{p,\min}$, a portion of the upper image is assumed to contain sky. The horizon line y_{horizon} is then estimated as: $y_{\text{horizon}} = (h)/2\left((\theta_p - \theta_{p,\min})/|\theta_{p,\min}|\right) + \delta$, where h is the image height, and δ is an offset for fine-tuning the vertical placement of the horizon. The ground projection is then recalculated using only the ground-visible region of the image, defined by the polygon: $\left[(0, y_{\text{horizon}}), (w, y_{\text{horizon}}), (w, h), (0, h)\right]$, where w and h denote the width and height of the image, respectively. Moreover, once $\theta_p > \theta_{p,\min}$, any additional area in the upper image corresponds to sky, regardless of further pitch increases. As a result, the distance from the image center to the top of the ground region (k_f) should remain constant. This distance is therefore set to a predefined constant. In this work, it is set to the maximum expected range of detection model $k_f^{\max}$.

In addition, homography transformation can suffer from distortion when applied to oblique images [25]. This limitation arises because the homography transformation does not accurately account for true depth variations in oblique scenes. To address this, vertical pixel scaling is applied to adjust vertical pixel coordinates before applying the transformation, using: $y' = y_{\text{horizon}} + ((y - y_{\text{horizon}})/(h - y_{\text{horizon}}))^{\gamma} \cdot (h - y_{\text{horizon}})$ where y is a vertical pixel coordinate, y' is the adjusted vertical pixel coordinate of y, and $\gamma > 1$ controls the degree of nonlinearity. The higher the γ, the higher the compression near the horizon and the higher the stretching near the bottom, enhancing the correction of depth variations in oblique views. This transformation improves geolocation accuracy near the horizon by reducing perspective distortion before applying the projection.

a) Calibration procedure: The horizon-aware ground projection correction and vertical pixel scaling methods were calibrated using the training data for the geolocation approximation task in the dataset described in Section V. This dataset contains several image sequences from the Thermal UAV platform. Each frame in every sequence includes the ground-truth geolocation of a designated target along with the corresponding UAV telemetry. The parameters of the geolocation approximation algorithm, including ccf, $\theta_{p,\min}$, δ, $k_f^{\max}$, and γ, were varied and systematically adjusted to minimize geolocation errors between the ground-truth and approximated geolocations. After calibration, the following parameters were used ccf $= 0.85$, $\theta_{p,\min} = -20$, $\delta = 5$, $k_f^{\max} = 3000$ (m), and $\gamma = 1.5$

C. RGB UAV Platform

The RGB UAV platform (**UAV-RGB**) is built on a drone [see Fig. 2(c)] that has been extensively customized by SKYLD Security and Defense, an ICT and consulting company specializing in Homeland Security. The customisation includes designing and 3D-printing specialized components, integrating multiple payloads, and implementing a tailored networking architecture. The drone is equipped with an onboard NVIDIA Jetson Orin 64 and two main sensors: RGB and thermal cameras. However, this platform only focuses on RGB imagery.

[4] https://docs.opencv.org/4.x/d1/de0/tutorial_py_feature_homography.html

The RGB camera streams video over the real-time streaming protocol (RTSP) through a dedicated onboard network established within the drone, using H26 radio communication to link the onboard systems with the ground station over TCP/IP. Each video stream is transmitted independently to the NVIDIA Jetson Orin 64 platform onboard the drone at 30 frames per second (FPS). The onboard Jetson processes the data and transmits detection, tracking, geolocation approximation results, annotated video streams, and telemetry data to the ground station. The camera operates at 30 FPS, whereas our onboard system can process image inputs at rates of up to 400 FPS, supporting multiple concurrent streams and advanced post-processing computations in real time. Overall, the RGB UAV platform provides high-frame-rate RGB imaging with real-time onboard processing via Jetson Orin, enabling flexible, robust detection and efficient data transmission.

1) Object Detection: A single-shot multibox detector (SSD) [26] with a MobileNetV2 [27] backbone is used for real-time object detection. This architecture offers a strong balance between speed and accuracy, making it particularly suitable for edge deployment on resource-constrained platforms such as the NVIDIA Jetson Orin. The SSD-MobileNetV2 model was initially pretrained on the COCO dataset and then retrained on a custom annotated dataset of RGB aerial imagery (see Section V), covering object classes such as persons, vehicles, vessels, and oil spills. Fine-tuning was performed in PyTorch using the following hyperparameters: 200 training epochs, batch size of 20, 12 data loader workers, and an initial learning rate of 0.01. The model was optimized using SGD with a momentum of 0.9 and a weight decay of 5e-4, with a Cosine Annealing learning rate scheduler over 100 epochs per cycle. The IoU threshold for the multibox loss function was set to 0.5. These settings ensured stable convergence and high detection accuracy on the aerial dataset. The model was then converted and optimized with NVIDIA TensorRT for real-time deployment on the Jetson Orin platform. This allowed for substantial inference acceleration while maintaining detection accuracy, achieving real-time performance of up to 400 FPS per stream. The model is integrated through NVIDIA's DetectNet libraries, part of the Jetson JetPack software suite, which provides a seamless pipeline for optimized inference. Notably, no explicit preprocessing is required at runtime, as the data pipeline is managed internally by the DetectNet framework, simplifying integration and maximizing throughput. All inference is GPU-accelerated, with system CPU usage remaining below 10%, even under continuous streaming and detection workloads.

2) Object Tracking: The platform uses the built-in tracking module of NVIDIA's DetectNet framework, which provides lightweight, real-time object association through intersection-over-union (IoU)-based matching. Although DetectNet supports Kanade–Lucas–Tomasi (optical flow) tracking [28], the IoU-based method was preferred for its simplicity, speed, and suitability for sea-based aerial data. This IoU-based tracker operates on bounding boxes produced by the SSD-MobileNetV2 detector. Each frame is analyzed to determine whether new detections sufficiently overlap with previously tracked objects, based on configurable IoU thresholds. This allows for consistent object identity across frames without the need for deep feature reidentification or motion prediction. Tracking is performed onboard the NVIDIA Jetson Orin, using GPU-accelerated inference pipelines that ensure minimal overhead. To determine suitable tracker parameters for maritime scenarios involving boats and persons, an iterative empirical tuning process was employed. Representative RGB UAV training sequences (see Section V) covering typical sea conditions, including moving boats of varying sizes and human subjects, were used in the tuning process. Based on this tuning, the tracker was configured with a minimum of three consecutive detections to validate a new track, a maximum of 15 missed detections before dropping a track, and an IoU threshold of 0.5 for box association. These values ensure robust tracking under the high background motion and small object sizes typical of aerial maritime imagery while maintaining real-time performance.

3) Object Geolocation Approximation: The RGB UAV platform employs the same geolocation approximation algorithm as used in the Therm UAV platform. However, to enhance precision and operational relevance, the system filters out detections estimated to be located more than 1 km away from the drone. These distant detections are excluded from the "object of interest" set to avoid error amplification and reduce irrelevant data. This filtering, combined with object tracking IDs, ensures that only nearby and relevant detections are geotagged and transmitted downstream for mission-critical decision-making.

D. Satellite-Based Systems

Satellite monitoring has an essential role in strengthening maritime surveillance, particularly in vessel monitoring. In this work, Sentinel-1 and Sentinel-2 images are used to achieve vessel detection. Sentinel-1 operates in a C-band SAR sensor in both ascending and descending orbits with a revisit time of six days and a spatial resolution of up to 10 m. Sentinel-2, which captures optical multispectral images, has an optimal spatial resolution of 10 m and, depending on the spectral band, it can have up to 60 m, while its revisiting time is five days. The advantages of Sentinel-1 are that it can record images regardless of weather conditions, day and night, making it extremely reliable for continuous monitoring of the sea. Its radar capabilities allow it to detect vessels even in cloudy or low visibility conditions [29]. On the other hand, Sentinel-2 [30], with its high spatial resolution, provides data that can be used to detect vessels, vessel waves, or oil spills. Although in cloudy or poor light conditions, its effectiveness is reduced. Another key advantage of these satellites over other commercial ones is that their data are freely available, which makes them an easy and cost-effective solution for large-scale marine monitoring. These satellite images offer a ground sample distance (GSD) of approximately 10 m. This allows for the detection of medium to large vessels, but limits the identification of smaller targets. In the case of Sentinel-2, where only RGB bands with a spatial resolution of 10 m were used, vessels must have a length and width of at least

10 m (corresponding to one pixel) to be reliably detected in the image. Objects smaller than the spatial resolution of the sensor cannot be distinguished in the images. However, these satellites are limited to recording data every five days, unlike other commercial satellites that record data every day or more frequently.

1) Object Detection: Recent advances in deep learning, particularly in CNN architectures, have significantly improved object detection. Early models such as R-CNN are not efficient due to their two-step processing, while modern one-stage models such as YOLO [31] achieve faster and more efficient predictions. Previous studies [32], [33] have demonstrated that YOLOv9 achieves higher accuracy and detection efficiency compared to older versions, requiring fewer parameters and lower computational cost. Based on these developments, this study adopts the YOLOv9 model for near-real-time vessel detection in satellite images for maritime surveillance purposes.

To ensure the suitability of YOLOv9, a basic comparison between the YOLOv7 (YOLOv7-E6E) and YOLOv9 (YOLOv9-E) models was performed using publicly available datasets: 1) a combination of Google Earth, Baidu Map [34], and PlanetScope [35] satellite imagery for optical imagery processing and 2) the SAR-Ship-Dataset [36] for SAR imagery processing. With these datasets, both YOLOv7 and YOLOv9 models were trained and tested under identical hyperparameter configurations to ensure a fair evaluation. On the SAR dataset, YOLOv7 achieved a precision of 0.894, while YOLOv9 demonstrated higher performance under the same conditions. For the optical imagery, YOLOv9 consistently outperformed YOLOv7, achieving a precision of 0.850 compared to 0.772. Based on these results, YOLOv9 was selected for object detection in the satellite-based systems.

As part of this work, two separate models were developed: one trained on optical images in the newly created dataset (see Section V) and the other on the SAR-Ship-Dataset. The models were trained to detect only one object class, namely "ship," and the detected objects were identified using bounding boxes, while the rest of the areas were considered as background.

Before applying detection algorithms on satellite images, both Sentinel-1 and Sentinel-2 satellite images underwent several preprocessing steps to ensure data quality during the next steps. For Sentinel-1 (SAR) imagery, preprocessing included subsetting to the area of interest, land masking to remove nonrelevant sea or land areas, thermal noise removal, radiometric calibration to convert pixel values into physically meaningful backscatter coefficients, speckle filtering to reduce inherent radar noise, applying precise orbit files for accurate geolocation, and finally, terrain correction to compensate for topographic distortions. These steps were implemented using the SnapPy library. In the case of Sentinel-2 (optical) data, Level-1 images were used, which already include basic pixel-value correction, orthorectification, and radiometric calibration. Additional preprocessing steps applied to these optical images involved spatial and spectral subsetting to focus on relevant bands and regions, as well as land masking to exclude unwanted areas.

As for training, the optical images were split into 1280 × 1280 pixel tiles. All tiles containing vessels were kept, while for the rest, only 1% was kept to be used for the training as background. The SAR images were cropped to 256 × 256 pixels for the model's training. During training, the batch size was chosen equal to 6, and hyperparameter validation was applied to 10% of the dataset. In addition, the models were run for 300 epochs, while the first 32 layers of the main YOLOv9-E model were frozen. Moreover, several data augmentation methods were used, such as HSV augmentation, image rotation, translation, scale, flip, shear, and perspective, which contributed to increasing the diversity of training data, enhancing the model's ability to detect vessels under different environmental conditions.

To evaluate the performance of both models, a check every 100 epochs was applied, allowing continuous optimization. This architecture ensures that this model is adequately responsive to large-scale open-sea region observation. For the model's training, the hardware systems used include the CPU AMD Ryzen 7 5800X 8-core processor, 32 GB of DDR4 3200MHz RAM, and an NVIDIA GeForce RTX 3090 with 24 GB of GPU. Since the model was trained on high-resolution images (Sentinel-2 with a pixel size of approximately 10 m), it is expected to perform effectively on datasets with similar spatial resolution, while its applicability to very high-resolution images (a pixel size of smaller than 2 m) may be limited.

E. Sensor Platform Integration

All sensor platforms are configured to synchronize their system clocks with a common network time protocol (NTP) server to ensure consistent timestamping across platforms. Results from each sensor, including object detections, tracking outputs, and estimated geolocations of the detected objects, are transmitted to the central system through a message queuing telemetry transport (MQTT) broker. This broker acts as a lightweight message-passing middleware, providing a reliable and scalable mechanism for real-time communication among sensor platforms. Each platform publishes its outputs to designated topics, while the central system subscribes to these topics to receive the incoming information from the sensors.

V. MAMMS Dataset

Due to the limited availability of datasets for maritime threat detection using multialtitude and multimodal sensors, this work introduces a novel dataset called MAMMS. This dataset includes data from ground-based sensors, low-altitude UAVs, and satellite-based systems. It is divided into two subsets.

1) *Ground-Based and Low-Altitude Sensors (MAMMS-GL):* Contains data from the AIT Smart Sensor platform, the RGB UAV platform, and the Thermal UAV platform, supporting object detection, object tracking, and geolocation approximation tasks.
2) *Optical Satellite (MAMMS-OS):* Contains optical (RGB) satellite imagery data, supporting the task of vessel detection.

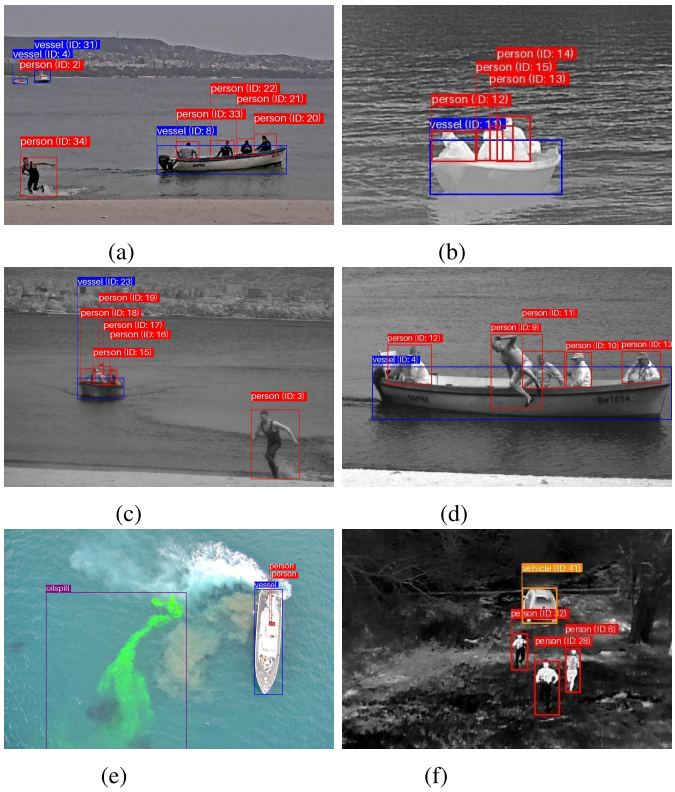


Fig. 4. Sample images from the MAMMS-GL dataset, showing images from different sensors, including RGB, thermal, UV, and SWIR AIT Smart Sensor platform, as well as RGB and thermal UAVs, with examples of four object classes: person, vessel, vehicle, and oil spill (a) GS-RGB. (b) GS-Therm. (c) GS-UV. (d) GS-SWIR. (e) UAV-RGB. (f) UAV-Thermal.

A. Ground-Based and Low-Altitude Sensors Dataset (MAMMS-GL)

This dataset comprises the data collected in simulated real-world scenarios using the AIT Smart Sensor, RGB UAV, and Thermal UAV platforms. The dataset can be divided into three parts for three different tasks: object detection, object tracking, and geolocation approximation. The following paragraphs describe each task-specific dataset within the MAMMS-GL dataset.

For the object detection task, the data collection was conducted in two different locations to cover a range of environmental conditions. The scenarios include various person, vessel, and vehicle movements and interactions between them, as well as simulated oil spills using an artificial, environmentally safe substance [see Fig. 4(e)] instead of real oil. Written informed consent was obtained from all actors in the scenarios. Annotations were made for all visible objects, whether fully or partially within the sensor's FOV, using axis-aligned bounding boxes. Each box was labeled with the object class (either person, vessel, vehicle, or oil spill). Examples of annotations are shown in Fig. 4. Tables I and II detailed the training and test sets.

The object-tracking dataset consists of GS-RGB, GS-UV, GS-SWIR, GS-Thermal, and UAV-Thermal sequences from the detection task, following the same training and testing split. Details on both training and test sets for object tracking are provided in Tables I and II. Track IDs are consistently preserved throughout each sequence, even when the objects were occluded or temporarily disappeared from the FOV, to support long-term tracking evaluation.

TABLE I
TRAINING DATA FROM THE MAMMS-GL DATASET FOR GROUND-BASED AND LOW-ALTITUDE OBJECT DETECTION AND TRACKING

Scenario	Sensor	#images	#persons	#vessels	#vehicles	#oilspills	#tracks
bg1	GS-RGB	378	3,141	769	76	-	24
	GS-SWIR	385	2,288	822	-	-	15
	GS-Therm	386	2,675	778	-	-	25
	GS-UV	386	2,479	897	-	-	15
bg3	GS-RGB	728	6,791	1,523	-	-	16
	GS-SWIR	800	6,693	1,846	-	-	14
	GS-Therm	742	6,183	1,485	-	-	19
	GS-UV	702	5,117	1,657	-	-	16
	UAV-RGB	245	2,580	595	29	-	-
bg4	GS-RGB	589	4,148	1,367	162	-	42
	GS-SWIR	754	6,591	2,091	13	-	32
	GS-Therm	635	4,145	1,471	-	-	26
	GS-UV	746	6,669	2,204	17	-	30
	UAV-RGB	53	545	312	0	-	-
bg5	GS-RGB	486	1,792	676	-	-	17
	GS-SWIR	583	1,751	1,165	-	-	9
	GS-Therm	499	450	861	-	-	12
	GS-UV	530	879	995	-	-	6
bg7	GS-RGB	756	2,758	1,009	69	-	24
	GS-Therm	810	1,165	1,869	-	-	13
	UAV-Therm	808	8,801	1,395	408	-	44
	UAV-RGB	623	4,943	2,328	0	-	-
bg9	GS-RGB	357	1,060	611	14	-	14
	GS-SWIR	386	581	1,094	-	-	7
	GS-Therm	391	1,037	685	-	-	8
	GS-UV	387	600	1,250	-	-	7
bg10	GS-RGB	459	1,798	1,051	144	-	17
	GS-SWIR	449	1,297	1,559	-	-	9
	GS-Therm	454	1,192	972	-	-	6
	GS-UV	387	986	1,340	-	-	11
	UAV-RGB	381	4,503	1,200	22	-	-
bg11	GS-RGB	391	5,203	1,255	-	-	30
	GS-SWIR	409	4,085	1,325	-	-	30
	GS-Therm	406	4,217	1,102	-	-	33
	GS-UV	410	3,771	1,356	-	-	30
	UAV-RGB	509	6,700	1,235	178	-	-
bg12	GS-RGB	263	776	958	-	-	9
	GS-SWIR	300	753	306	-	-	7
	GS-Therm	297	652	36	-	-	5
	GS-UV	301	790	416	-	-	7
bg13	UAV-RGB	322	3,453	929	65	-	-
cy1	UAV-RGB	970	2,115	399	-	939	-
	UAV-Therm	969	1,795	906	-	-	5
cy2	UAV-RGB	1,204	1,569	888	-	1201	-
	UAV-Therm	1,193	3,962	2,151	-	-	22
Total		24,219	135,479	51,139	1,197	2,140	656

For the geolocation approximation task, 13 thermal UAV sequences were collected. Such sequences cover both parameter-constrained scenarios and unconstrained scenarios. The parameter-constrained scenarios were designed with specific variations in three parameters: UAV ground projection position, altitude, and pitch angle, as shown in Table III. The unconstrained scenarios correspond to the previously described data collection settings for object detection and tracking. In these scenarios, the UAV telemetry values were not restricted during data collection. Across all parameter-constrained and unconstrained scenarios, each sequence contains a series of images capturing a specific target object (either a vessel or a person). Each image includes a bounding box around the target object, the target's GNSS ground-truth coordinates (from a GNSS tracker), and UAV telemetry data (e.g., focal length, zoom level, position, altitude, gimbal angles, and timestamp). This enables frame-by-frame geolocation estimates to be compared with ground truth. Details of all training and test sequences for this task,

TABLE II
TEST DATA FROM THE MAMMS-GL DATASET FOR GROUND-BASED AND LOW-ALTITUDE OBJECT DETECTION AND TRACKING

Scenario	Sensor	#images	#persons	#vessels	#vehicles	#oil spills	#tracks
bg2	GS-RGB	871	8,233	2,506	-	-	26
	GS-SWIR	846	6,069	2,197	-	-	14
	GS-Therm	845	5,679	1,723	-	-	20
	GS-UV	845	5,131	2,318	-	-	18
	UAV-Therm	842	6,780	1,947	-	-	17
	UAV-RGB	802	4,021	3,507	0	0	-
bg6	GS-RGB	628	3,536	1,485	133	-	28
	GS-SWIR	569	3,531	2,116	-	-	14
	GS-Therm	565	3,851	1,823	-	-	15
	GS-UV	571	3,818	2,038	-	-	16
	UAV-RGB	115	1,004	614	0	0	-
bg8	GS-RGB	1,526	7,434	5,345	-	-	24
	GS-SWIR	1,515	5,093	4,609	-	-	25
	GS-Therm	1,515	6,337	4,046	-	-	22
	UAV-Therm	1,513	9,455	7,426	425	-	47
	UAV-RGB	1,739	12,754	3,259	0	0	-
cy3	UAV-RGB	1,072	440	1,552	-	799	-
	UAV-Therm	1,059	1,714	1,840	-	-	9
Total		17,438	94,880	50,351	558	799	295

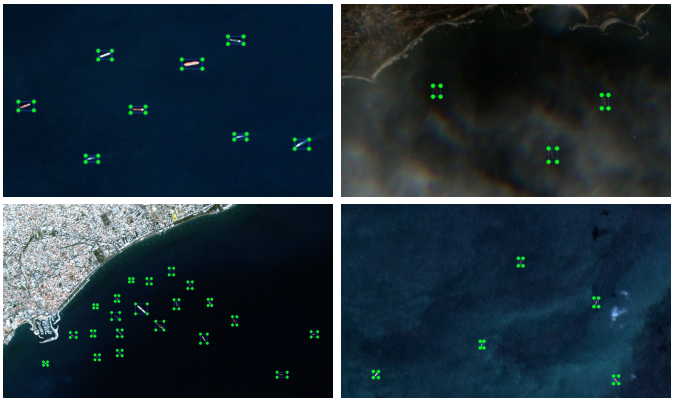


Fig. 5. Samples from the MAMMS-OS dataset, showing RGB Sentinel-2 images with "ship" annotation.

including the number of images and GNSS ground-truth coordinates available in each sequence, are provided in Tables IV and V, respectively. Eight sequences were used for calibration and hyperparameter tuning; the remaining five were used for testing. Sequences "rd1"–"rd7" are from parameter-constrained scenarios, while sequences "bg7," "cy1," "cy2," and "cy4"–"cy6" are unconstrained scenarios.

It is also worth noting that a subset of this MAMMS-GL has been part of the PETS 2025 Challenge [37] and is publicly available at http://pets2025.net

B. Optical Satellite Dataset (MAMMS-OS)

This dataset contains Sentinel-2 images obtained from the Copernicus website [38] for two areas of interest in Cyprus and Bulgaria. For these areas, all available images were collected for nine months in total. Thus, 111 satellite images were annotated, and all vessels within them were classified as "ship." An important advantage of this dataset is that it includes images in different light and weather conditions, supporting the model to detect vessels even in unclear sky conditions (see Fig. 5). Only the RGB bands of the satellite images were used for vessel detection, as these bands offer the highest spatial resolution of 10 m in Sentinel-2 images. The 10-m spatial resolution allows for the identification of vessels of comparable or larger size, while lower resolution bands (20 or 60 m) would not allow for reliable detection of smaller vessels. From the data annotation procedure, 3708 vessels in total were detected. To proceed with the model's implementation, the dataset was split into training, testing, and validation sets in percentages of 80%–10%–10%, respectively. In particular, the objects used for training were 2929 "ships," while for the testing and validation were 365 and 414 "ships," respectively.

VI. EXPERIMENTAL SETUP

A. Experimental Datasets

Extensive experiments were conducted to evaluate and benchmark the performance of the proposed sensor platforms across tasks relevant to each platform. The experiments were performed on the proposed benchmarking dataset MAMMS (see Section V) and on an external dataset for completeness. Overall, three datasets were used for evaluation and benchmarking.

1) MAMMS-GL dataset for object detection, tracking, and geolocation approximation on the AIT Smart Sensor and the two UAV platforms. The details of the training and testing splits are provided in Section V.
2) MAMMS-OS dataset for object detection on optical (RGB) satellite imagery in the satellite-based systems. The data split is described in Section V.
3) SAR-Ship-Dataset [36] for object detection on SAR satellite imagery in the satellite-based systems. The SAR-Ship-Dataset is an existing open dataset, containing images from the Chinese Gaofen-3 and Sentinel-1 satellites. It comprises 102 Gaofen-3 images and 108 Sentinel-1 images. These images were cropped, resulting in 39,729 image chips. From this dataset, 80% was used for training, 10% for testing, and 10% for validation.

B. Evaluation Methods

1) Object Detection Evaluation: To benchmark object detection performance on the AIT Smart Sensor (GS-RGB, GS-UV, GS-SWIR, and GS-Therm), RGB UAV (UAV-RGB), and Thermal UAV platforms (UAV-Therm), 5 state-of-the-art models were selected to compare with object detectors developed for each sensor. These models are as follows.

1) *YOLOv5* [39], *YOLOX* [40], *YOLOv8* [41], and *YOLOv11* [23]: Four pretrained models in the YOLO family were used as baselines, particularly the *YOLOv5s*, *YOLOv8s*, *YOLOX-l*, and *YOLO11s* variants. For all of these models, inference was performed using the following settings: 640 × 640 input size, confidence threshold 0.25, IoU threshold 0.45; and the maximum number of detections 300.
2) *RTDETR* [42]: This model is real-time detection transformer. The pretrained *RT-DETR Large* was used without fine-tuning. Inference was performed with the same parameters as the YOLO baselines.
3) *DET-GS*: the proposed detection model for GS-RGB, GS-UV, GS-SWIR, and GS-Therm. The training hyperparameters are described in Section IV-A1. For inference, the input size was set to 640 × 640, and the

TABLE III
DESCRIPTIONS OF PARAMETER-CONSTRAINED SCENARIOS USED TO COLLECT DATA FOR THE UAV GEOLOCATION APPROXIMATION TASK IN THE MAMMS-GL DATASET

Scenario	Ground Projection Position	Altitude	Camera Pitch
rd1: Forward-moving drone with fixed altitude and fixed camera pitch	Varied (moving towards the target)	Fixed at 10 m	Fixed at -25°
rd2: Forward-moving drone with fixed altitude and fixed camera pitch	Varied (moving towards the target)	Fixed at 10 m	Fixed at -25°
rd3: Hovering drone with fixed altitude and varying camera pitch	Fixed (hovering)	Fixed at 15 m	Varied (-50° to -26°)
rd4: Ascending drone with varying altitude and varying camera pitch	Fixed (ascending)	Varied (10 to 15 m)	Varied (-45° to -30°)
rd5: Forward-moving drone with fixed altitude and varying camera pitch	Varied (moving towards the target)	Fixed at 8 m	Varied (-64° to -18°)
rd6: Forward-moving drone with fixed altitude and varying camera pitch	Varied (moving towards the target)	Fixed at 10 m	Varied (-90° to -19°)
rd7: Forward-moving drone with fixed altitude and varying camera pitch	Varied (moving towards the target)	Fixed at 6 m	Varied (-40° to -0°)

TABLE IV
TRAINING DATA FROM THE MAMMS-GL DATASET FOR THE GEOLOCATION APPROXIMATION TASK ON GROUND-BASED AND LOW-ALTITUDE SENSORS

Scenario	#images	#GNSS data points
cy1	969	906
cy2	1,193	1,134
rd1	514	415
rd2	588	527
rd3	1,388	1,388
rd4	322	322
rd5	395	337
rd6	589	529
Total	5,958	5,558

TABLE V
TEST DATA FROM THE MAMMS-GL DATASET FOR THE GEOLOCATION APPROXIMATION TASK ON GROUND-BASED AND LOW-ALTITUDE SENSORS

Scenario	#images	#GNSS data points
bg7	808	697
cy4	351	343
cy5	225	225
cy6	201	201
rd7	561	464
Total	2,146	1,930

confidence threshold was set to 0.3. All other parameters were kept the same as in the YOLOX baseline.

4) *DET-UAVT*: the proposed detection model for UAV-Therm. The training hyperparameters are provided in Section IV-B1. For inference, the same hyperparameters as those used in the YOLO baselines were applied.
5) *DET-UAVR*: the proposed detection model for UAV-RGB. Details on the training process can be found in Section IV-C1. An input size of 3840×2160 pixels and a confidence threshold of 0.5 were used. The maximum number of detections was not limited.

Note that for GS-RGB and GS-UV, only vessel detection was evaluated; persons and vehicles were ignored. This is because, in real-world operation, both sensors were intended to focus solely on vessel detection. Moreover, for GS-RGB, all models were tested on only two ("bg2" and "bg6") of the three available test sequences due to computational resource limitations.

For ground-based and low-altitude sensor platforms, the number of true positives (TPs), false positives (FPs), false negatives (FNs), false negative rate (FNR), precision, recall, F1-score, and average precision (AP) were used. The IoU threshold was set to 0.5.

As for the satellite-based systems, no baseline comparisons were performed due to the absence of suitable open-source models for vessel detection in optical and SAR satellite imagery. The evaluation was therefore conducted on two object detection models: the proposed YOLOv9 models, one trained on Sentinel-2 optical (RGB) images and the other on Sentinel-1 SAR images. The training details are described in Section IV-D1. For inference, the confidence threshold was set to 0.01 for optical images and 0.1 for SAR images. The IoU threshold was fixed at 0.45, and the number of detections was capped at 1,000. Regarding the evaluation metrics, the accuracy, precision, recall, mAP@0.5, and mAP@0.5:0.95 metrics were used.

2) Object Tracking Evaluation: Three state-of-the-art tracking methods are evaluated alongside the proposed trackers developed for each sensor.

1) *DeepSORT [43]:* The DeepSort tracker was configured with a maximum overlap distance of 0.7 for matching detections to existing tracks and retained tracks for up to 30 frames without updates (track buffer). The number of frames that a track remains in the initialization phase was set to 3.
2) *ByteTrack [44]:* The ByteTrack tracker was configured with a high-confidence threshold of 0.25 and a low-confidence threshold of 0.1. New tracks were initiated when detection confidence reached 0.25, and the track buffer was 30 frames. The association similarity threshold was set to 0.8.
3) *BoT-SORT [24]:* The BoT-SORT tracker used the same high-confidence threshold, low-confidence threshold, new-track threshold, track buffer, and matching threshold as in ByteTrack. In addition, BoT-SORT applied sparse optical flow for global motion compensation and set the proximity and appearance thresholds for identity matching to 0.5 and 0.8.
4) *TRK-GS:* The proposed tracker for GS-SWIR and GS-Therm. The hyperparameters used are described in Section IV-A2.
5) *TRK-UAVT:* The proposed tracker for UAV-Therm). The hyperparameters used are described in Section IV-B2.

Each of the baseline tracking methods is paired with all five detection models: YOLOv5, YOLOX, YOLOv8, YOLOv11, and RTDETR for comprehensive comparisons.

The metrics used in MOTChallenge [45] are adopted to assess both target localization accuracy and target redetection capability. These include ID switches (IDSWs), ID F1-Score (IDF1) [46], multiple object tracking accuracy (MOTA) [45], multiple object tracking precision (MOTP) [45], and higher order tracking accuracy (HOTA) [47]. The IoU threshold was set to 0.5.

3) Geolocation Approximation Evaluation: For this task, the evaluation was conducted using the test data described in Section V-A. As test data for geolocation estimation is only available for UAV-Therm, the evaluation was conducted solely on this platform. The proposed geolocation approximation method for UAV-Therm was compared against its variations as follows.

1) *GEO-UAVT:* The proposed method without horizon-aware ground projection correction and without vertical pixel scaling for geolocation stabilization
2) *GEO-UAVT+:* The proposed method with horizon-aware ground projection correction but without vertical pixel scaling for Geolocation Stabilization
3) *GEO-UAVT++:* The complete proposed method with both horizon-aware ground projection correction and vertical pixel scaling for geolocation stabilization

The same method as [48] was adopted to compare the performance of each method. The Haversine distance between the estimated and ground-truth geolocations in the test set was measured and treated as the distance error. The minimum, maximum, mean absolute error (MAE), and root mean square error (RMSE) of the computed distance errors were then calculated to compare the performance of each method.

VII. Results and Discussion

A. Object Detection Results

1) Ground-Based and Low-Altitude Sensors: Table VI shows the object detection benchmarking results on the MAMMS-GL dataset for all ground-based (GS-RGB, GS-UV, GS-SWIR, and GS-Therm) and low-altitude sensors (UAV-RGB and UAV-Therm).

Across **GS-RGB, GS-UV, GS-SWIR, and GS-Therm**, DET-GS outperformed the YOLO baselines in both vessel and person detection in most cases. The only exceptions were vessel detection on GS-SWIR, where YOLOv11 slightly outperformed DET-GS, and person detection on GS-SWIR and GS-Therm, where YOLOX performed better than DET-GS . This indicates that the proposed DET-GS is effective for vessel detection, which is the main focus of the platform, whereas person detection still requires improvement, although it is not the primary priority.

Compared to RTDETR, although DET-GS performed worse overall, the results reveal distinctive advantages of DET-GS in vessel detection. In particular, on average across all ground-based sensors, DET-GS achieved a substantial 65.2% reduction in FP with decreases of 40.5% in AP and 24.6% in F1-score relative to RTDETR. This demonstrates that DET-GS generated significantly fewer false detections, outweighing the moderate decreases in overall accuracy metrics. Although some real-world surveillance systems prioritize minimizing FNs to maximize detections, this often increases false alarms. Reducing false alarms is crucial because excessive alerts can cause unnecessary resource deployment and disrupt multi-sensor fusion and tracking processes that depend on reliable detections [49]. Therefore, in scenarios where minimizing false alarms is a priority, DET-GS can be a more suitable choice.

TABLE VI
Object Detection Benchmarking Results on the MAMMS-GL Dataset for Ground-Based and Low-Altitude Sensors

Class	Model	TP	FP	FN	FNR	Prec.	Recall	F1	AP
GS-RGB									
vessel	YOLOv5	847	49	3144	0.788	0.945	0.212	0.347	0.205
	YOLOX	1186	109	2805	0.703	0.916	0.297	0.449	0.283
	YOLOv8	823	**38**	3168	0.794	**0.956**	0.206	0.339	0.203
	YOLOv11	1086	105	2905	0.728	0.912	0.272	0.419	0.264
	RTDETR	**2923**	791	**1068**	**0.268**	0.787	**0.732**	**0.759**	**0.707**
	DET-GS	1662	274	2329	0.584	0.858	0.416	0.561	0.394
GS-UV									
vessel	YOLOv5	686	**8**	3670	0.843	0.988	0.157	0.272	0.157
	YOLOX	1126	9	3230	0.742	**0.992**	0.258	0.410	0.258
	YOLOv8	797	25	3559	0.817	0.970	0.183	0.308	0.182
	YOLOv11	1216	25	3140	0.721	0.980	0.279	0.435	0.277
	RTDETR	**2352**	303	**2004**	**0.460**	0.886	**0.540**	**0.671**	**0.526**
	DET-GS	1227	21	3129	0.718	0.983	0.282	0.438	0.281
GS-SWIR									
person	YOLOv5	1653	**14**	13040	0.887	**0.992**	0.113	0.202	0.112
	YOLOX	2089	58	12604	0.858	0.973	0.142	0.248	0.142
	YOLOv8	1741	**14**	12952	0.882	**0.992**	0.118	0.212	0.118
	YOLOv11	1845	40	12848	0.874	0.979	0.126	0.223	0.125
	RTDETR	**2553**	350	**12140**	**0.826**	0.879	**0.174**	**0.290**	**0.171**
	DET-GS	2063	391	12630	0.860	0.841	0.140	0.241	0.122
vessel	YOLOv5	2551	208	6371	0.714	0.925	0.286	0.437	0.281
	YOLOX	3217	**30**	5705	0.639	**0.991**	0.361	0.529	0.360
	YOLOv8	2777	78	6145	0.689	0.973	0.311	0.472	0.309
	YOLOv11	3519	211	5403	0.606	0.943	0.394	0.556	0.387
	RTDETR	**6718**	1204	**2204**	**0.247**	0.848	**0.753**	**0.798**	**0.734**
	DET-GS	3468	184	5454	0.611	0.950	0.389	0.552	0.379
GS-Therm									
person	YOLOv5	2570	64	13297	0.838	0.976	0.162	0.278	0.161
	YOLOX	3000	240	12867	0.811	0.926	0.189	0.314	0.184
	YOLOv8	2364	68	13503	0.851	0.972	0.149	0.258	0.149
	YOLOv11	2288	**47**	13579	0.856	**0.980**	0.144	0.251	0.144
	RTDETR	**4582**	597	**11285**	**0.711**	0.885	**0.289**	**0.435**	**0.283**
	DET-GS	2399	576	13468	0.849	0.806	0.151	0.255	0.132
vessel	YOLOv5	2938	**274**	4654	0.613	**0.915**	0.387	0.544	0.372
	YOLOX	4938	801	2654	0.350	0.860	0.650	0.741	0.604
	YOLOv8	2864	290	4728	0.623	0.908	0.377	0.533	0.363
	YOLOv11	4579	708	3013	0.397	0.866	0.603	0.711	0.562
	RTDETR	**6646**	2251	**946**	**0.125**	0.747	**0.875**	**0.806**	**0.830**
	DET-GS	5692	1854	1900	0.250	0.754	0.750	0.752	0.640
UAV-Therm									
person	YOLOv5	1127	11	16822	0.937	0.990	0.063	0.118	0.063
	YOLOX	1491	39	16458	0.917	0.975	0.083	0.153	0.082
	YOLOv8	536	5	17413	0.970	0.991	0.030	0.058	0.030
	YOLOv11	487	**3**	17462	0.973	**0.994**	0.027	0.053	0.027
	RTDETR	2377	364	15572	0.868	0.867	0.132	0.230	0.126
	DET-UAVT	**6982**	1534	**10967**	**0.611**	0.820	**0.389**	**0.528**	**0.366**
vessel	YOLOv5	2444	**36**	8769	0.782	**0.985**	0.218	0.357	0.217
	YOLOX	4072	162	7141	0.637	0.962	0.363	0.527	0.358
	YOLOv8	2520	41	8693	0.775	0.984	0.225	0.366	0.224
	YOLOv11	3947	80	7266	0.648	0.980	0.352	0.518	0.350
	RTDETR	**6907**	328	**4306**	**0.384**	0.955	**0.616**	**0.749**	**0.613**
	DET-UAVT	4816	369	6397	0.570	0.929	0.430	0.587	0.425
vehicle	YOLOv5	70	14	355	0.835	0.833	0.165	0.275	0.164
	YOLOX	101	40	324	0.762	0.716	0.238	0.357	0.198
	YOLOv8	86	9	339	0.798	0.905	0.202	0.331	0.200
	YOLOv11	68	**3**	357	0.840	**0.958**	0.160	0.274	0.160
	RTDETR	156	81	269	0.633	0.658	0.367	0.471	0.309
	DET-UAVT	**282**	23	**143**	**0.336**	0.925	**0.664**	**0.773**	**0.657**
UAV-RGB									
person	YOLOV5	3327	139	14892	0.817	0.960	0.183	0.307	0.181
	YOLOX	5015	287	13204	0.725	0.946	0.275	0.426	0.271
	YOLOV8	3453	**74**	14766	0.810	**0.979**	0.190	0.318	0.188
	YOLOV11	5170	449	13049	0.716	0.920	0.284	0.434	0.277
	RTDETR	8595	2164	9624	0.528	0.799	0.472	0.593	0.453
	DET-UAVR	**11458**	8371	**6761**	**0.371**	0.578	**0.629**	**0.602**	**0.516**
vessel	YOLOV5	3085	**363**	5847	0.655	**0.895**	0.345	0.498	0.313
	YOLOX	3311	1586	5621	0.629	0.676	0.371	0.479	0.304
	YOLOV8	3055	685	5877	0.658	0.817	0.342	0.482	0.313
	YOLOV11	3515	866	5417	0.606	0.802	0.394	0.528	0.371
	RTDETR	5254	3168	3678	0.412	0.624	0.588	0.606	0.515
	DET-UAVR	**6272**	2969	**2660**	**0.298**	0.679	**0.702**	**0.690**	**0.654**
oil spill	DET-UAVR	**709**	**10**	**90**	**0.113**	**0.986**	**0.887**	**0.934**	**0.887**

TABLE VII
OBJECT DETECTION BENCHMARKING RESULTS ON THE MAMMS-OS DATASET (OPTICAL) AND THE SAR-SHIP-DATASET (SAR) FOR THE SATELLITE-BASED SYSTEMS

Dataset	Accuracy	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Optical	0.70	0.735	0.652	0.692	0.260
SAR	0.95	0.923	0.837	0.922	0.619

For UAV-Therm, DET-UAVT consistently outperformed its baseline counterpart YOLOv11 across all classes. However, it produced more FPs than the original YOLOv11, resulting in lower precision for person and vessel detection. Despite this, DET-UAVT generally achieved better performance than all YOLO-based models in terms of F1-score and AP. Compared to the strong baseline RTDETR, DET-UAVT showed superior overall performance in person and vehicle detection. For vessel detection, however, RTDETR achieved the best results, with DET-UAVT ranking second among all models. This highlights both the strong performance of RTDETR and the competitive capability of DET-UAVT in vessel detection.

For **UAV-RGB**, the proposed DET-UAVR achieved the best overall performance in person and vessel detection, outperforming all baselines. RTDETR generally ranked second. Although no baseline is available for oil spill detection, DET-UAVR demonstrated strong results, with an F1 score of 0.934 and an AP of 0.887, highlighting its effectiveness for this task.

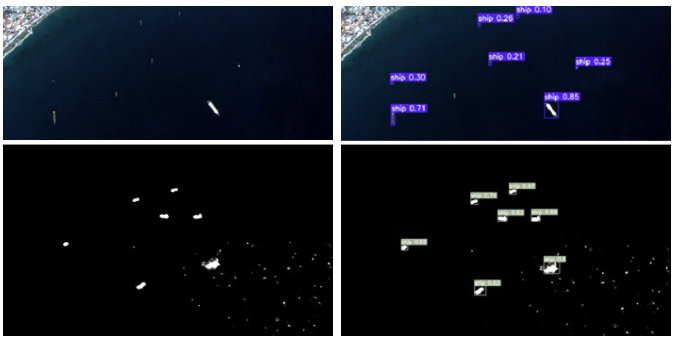
2) Satellite Sensors: The object detection performance of the satellite-based system on the **MAMMS-OS** dataset and the **SAR-Ship-Dataset** is presented in Table VII. For the **MAMMS-OS** dataset, the model achieved an accuracy of 0.70, a precision of 0.735, and a recall of 0.652, suggesting a moderate ability to correctly detect the vessels. A mAP@0.5 of 0.692 indicates accurate detection results, while a mAP@0.5:0.95 of 0.260 means that at tighter thresholds, the model shows lower accuracy. These results highlight the difficulties in detecting vessels in high-resolution optical images, as weather conditions, such as clouds, can affect the results.

On the other hand, the results on the **SAR-Ship-Dataset** indicate the high performance of the YOLOv9 model in detecting ships in SAR images with an accuracy of 0.95. Precision was found to be equal to 0.923, indicating that the majority of detections were correct, and recall was found to be equal to 0.837, meaning that the model successfully detected most of the existing ships. The average accuracy mAP@50 reached 0.922, highlighting the excellent performance of the model, while mAP@0.5:0.95 was equal to 0.619, meaning that the model is also accurate in more strict conditions of evaluation. The detection results of SAR images indicate the ability of the model to have a high performance in maritime surveillance using SAR imagery.

Figs. 6 and 7 present the detection results on the Sentinel-2 and Sentinel-1 images in the testing phase. The figures show the ground truth of the vessels on Figs. 6(a) and 7(a), while Figs. 6(b) and 7(b) indicate the detection results of the models.

(a) (b)

Fig. 6. Detection results on the MAMMS-OS dataset. (a) Ground-truth. (b) Models detections.



(a) (b)

Fig. 7. Detection results on the SAR-Ship-Dataset. (a) Ground-truth. (b) Models detections.

B. Tracking Results

Fig. 8 shows the object tracking benchmarking results on the **MAMMS-GL** dataset for ground-based and low-altitude sensors. For **GS-RGB**, TRK-GS, RTDETR+ByteTrack, and RTDETR+BoT-SORT achieved comparable MOTA scores, significantly outperforming other baselines. However, TRK-GS showed lower IDF1 compared to the other two, though still surpassing the remaining baselines. The reduced IDF1 also contributed to a lower HOTA score. Nevertheless, TRK-GS showed superior performance with significantly fewer IDSWs compared to trackers with higher IDF1. This demonstrates the strong performance of TRK-GS in maintaining accurate ID associations and tracking continuity. MOTP scores were similar across all methods, indicating comparable localization accuracy. Overall, the results indicate that, for GS-RGB, TRK-GS is effective at maintaining long tracks but may miss some due to FNs in detection or the exclusion of low-confidence tracks.

For **GS-UV and GS-SWIR**, the results differed from those for GS-RGB. TRK-GS achieved high IDF1 scores, comparable to top-performing methods, including RTDETR+ByteTrack and RTDETR+BoT-SORT. However, it exhibited lower MOTA than these two methods. This suggests strong identity preservation but poorer detection accuracy with higher FP and FN (see Table VI). Another strong method is YOLOX+BoT-SORT, which achieved higher IDF1 than TRK-GS for GS-UV. This is likely because YOLOX+BoT-SORT generated substantially fewer detections, including fewer FPs, than TRK-GS .

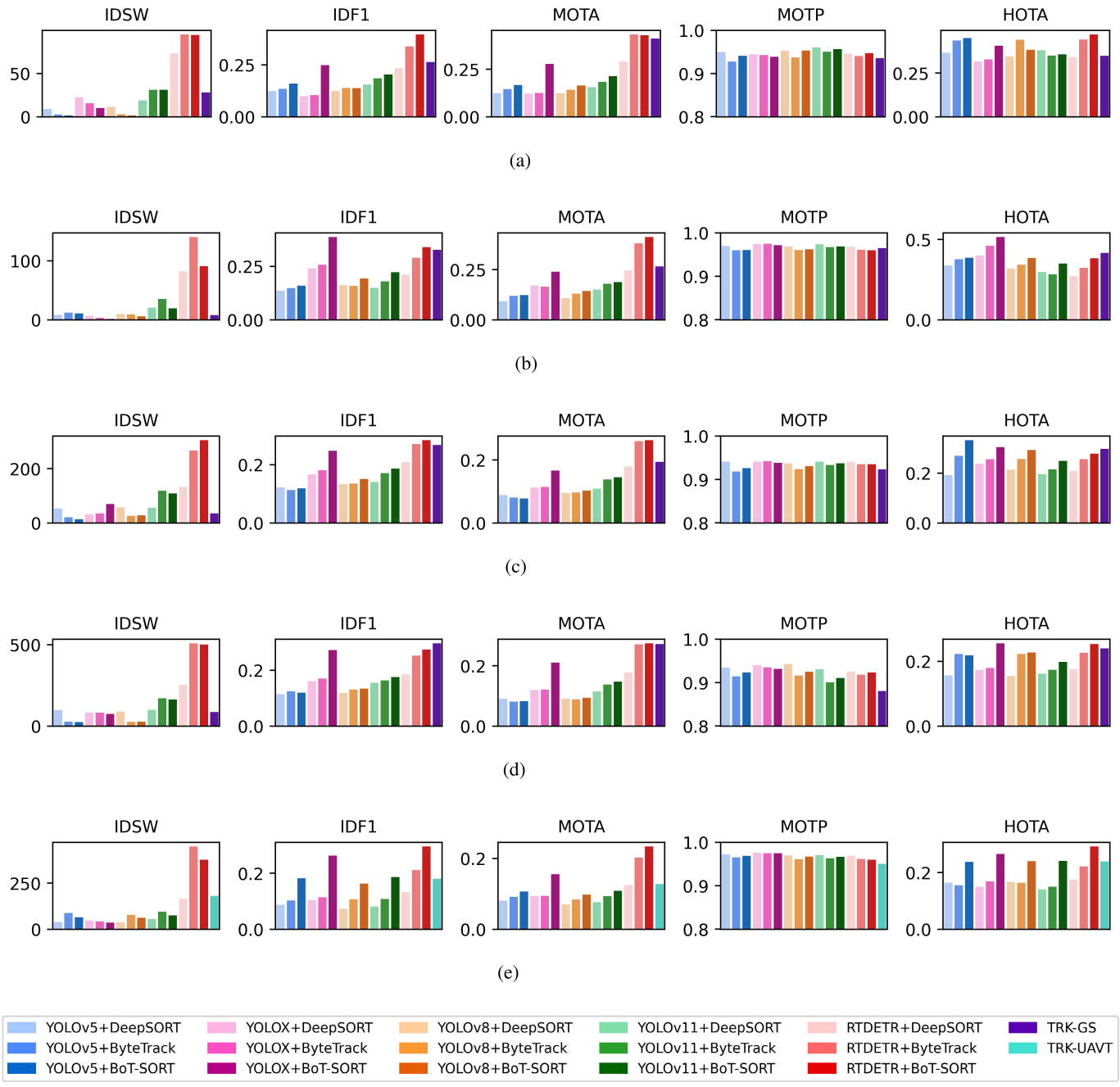


Fig. 8. Object tracking benchmarking results on the MAMMS-GL dataset for ground-based and low-altitude sensors. (a) GS-RGB. (b) GS-UV. (c) GS-SWIR. (d) GS-Therm. (e) UAV-Therm.

This lowered the chance of assigning IDs to incorrect objects or FPs, which in turn increased IDF1. This also affected HOTA, making TRK-GS underperform YOLOX+BoT-SORT in terms of HOTA. The IDSW results also indicate the strong identity preservation of TRK-GS, as its IDSW is lower than most baselines.

For **GS-Therm**, TRK-GS outperformed all other methods in IDF1 and achieved a relatively high MOTA score, indicating strong performance in consistent identity preservation and overall tracking accuracy. RTDETR+BoT-SORT and RTDETR+ByteTrack also performed well in both IDF1 and MOTA. However, TRK-GS demonstrated stronger performance in minimizing IDSW. This suggests that it maintained more consistent tracks than BoT-SORT and ByteTrack. As in the previous sensors, all methods showed similar performance in MOTP and HOTA.

For **UAV-Therm**, RTDETR+BoT-SORT achieved the highest scores in IDF1, MOTA, and HOTA. Close behind are ByteTrack+BoT-SORT and ByteTrack+RTDETR, and TRK-UAVT, which also showed competitive results across these metrics. Although RTDETR+BoT-SORT achieved the best overall performance, it exhibited substantially higher IDSW than TRK-UAVT. This suggests that TRK-UAVT maintains superior identity consistency, despite its lower IDF1, MOTA,

TABLE VIII
HAVERSINE DISTANCE ERRORS BETWEEN GROUND-TRUTH AND PREDICTED OBJECT GEOLOCATIONS ON THE MAMMS-GL DATASET

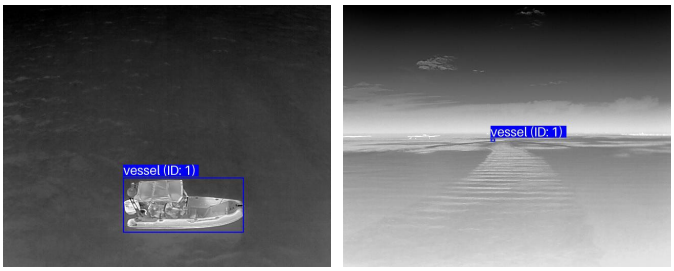
Scenario	Method	Min (m)	Max (m)	MAE (m)	RMSE (m)
bg7	GEO-UAVT	11.29	181.76	62.09	85.10
	GEO-UAVT+	10.63	131.71	49.70	65.56
	GEO-UAVT++	**9.03**	**77.86**	**33.37**	**40.56**
cy4	GEO-UAVT	**2.16**	**19.73**	5.74	6.63
	GEO-UAVT+	**2.16**	**19.73**	**5.64**	**6.55**
	GEO-UAVT++	**2.16**	**19.73**	5.80	6.62
cy5	GEO-UAVT	653.65	751.42	730.88	731.32
	GEO-UAVT+	365.74	455.83	427.82	428.35
	GEO-UAVT++	**20.20**	**75.08**	**42.93**	**44.83**
cy6	GEO-UAVT	19.68	69.05	46.48	49.69
	GEO-UAVT+	17.90	50.54	34.72	36.56
	GEO-UAVT++	**15.48**	**29.14**	**21.57**	**21.99**
rd7	GEO-UAVT	**4.52**	20.94	14.08	15.11
	GEO-UAVT+	**4.52**	18.68	12.71	13.46
	GEO-UAVT++	**4.52**	**16.33**	**9.79**	**10.42**

and HOTA scores. The lower IDF1, MOTA, and HOTA scores can be attributed to reduced vessel detection accuracy (see Table VI), as these metrics are influenced by detection accuracy. For MOTP, all methods achieved similarly high scores, suggesting that localization precision is consistent across all methods.

Overall, the relatively low scores across all metrics suggest inherent limitations of image-based tracking in this context. Relying on short-term appearance cues and spatial proximity, these methods struggle with appearance changes, reentries, and prolonged occlusions. These are common challenges in our data due to constant sensor motion.

C. Geolocation Approximation Results

Table VIII shows the Haversine distance errors between the ground-truth and predicted object geolocations on the UAV-Therm data from the **MAMMS-GL** dataset. In scenarios "bg7," "cy6," and "rd7," as the method progressed from GEO-UAVT to GEO-UAVT+ and finally to GEO-UAVT++, both MAE and RMSE values decreased, indicating enhanced geolocation accuracy at each correction step. In scenario "cy4," all three methods performed similarly. This is because most images in this scenario were captured with low pitch angles and without a visible sky area, as illustrated in Fig. 9(a). Therefore, the proposed horizon-aware ground projection correction and vertical pixel scaling for geolocation stabilization were not fully utilized. Scenario "cy5" clearly shows the performance of the proposed method. In this scenario, the vessel was captured from a great distance with high pitch angles (close to zero), as illustrated in Fig. 9(b), making geolocation approximation particularly challenging. GEO-UAVT produced an extremely high MAE (730.88 m) and RMSE (731.32 m). The application of horizon-aware correction in GEO-UAVT+ substantially reduced these errors. GEO-UAVT++ made further significant improvements, reducing the MAE to 42.93 m and RMSE to 44.83 m. This underscores the critical role of the vertical scaling step in scenarios where the pitch angle causes severe distortion near the horizon.



(a) (b)

Fig. 9. Examples of the majority of images in scenarios (a) "cy4" (where the vessel was captured at a low pitch angle without the sky in view) and (b) "cy5" (where the vessel was captured at a significant distance).

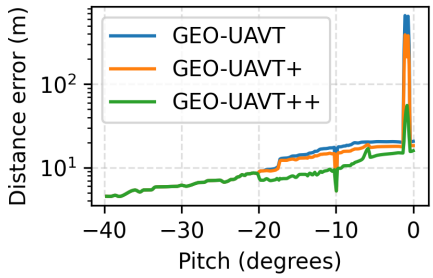


Fig. 10. Effect of UAV pitch angle on Haversine distance error between the ground-truth coordinates and the estimated coordinates from each method.

Fig. 10 illustrates the effect of UAV pitch angle on the Haversine distance error between the ground-truth coordinates and the estimated coordinates produced by each method. Across all methods, the errors clearly increased as pitch angles increased. The errors also escalated rapidly as the pitch angles approached 0. This emphasizes the effect of pitch angle on the approximation algorithms, particularly when it is close to 0 (nearly horizontal to the ground), and the captured image includes a portion of the sky. Comparing all three methods, for pitch angles less than $-20°$, the errors from all methods were equivalent, suggesting no impact from the horizon-aware ground projection correction and vertical pixel scaling. However, when the pitch angle exceeded $-20°$, the differences between the methods became apparent. GEO-UAVT++ showed the lowest errors, followed by GEO-UAVT+ and GEO-UAVT, respectively. This indicates the effectiveness of the proposed components in mitigating the effect of pitch angle in the original geolocation approximation algorithm. Moreover, GEO-UAVT++ produced errors lower than 10 m when the pitch angle was below $-10°$. For pitch angles higher than this, the error increased beyond 10 m and reached up to approximately 45 m. This suggests that lower pitch angles, particularly below $-10°$, are more effective for accurate geolocation.

D. Case Study

After evaluating the performance of individual sensor platforms across different tasks, a case study was conducted. The goal is to assess the effectiveness of the proposed sensor platforms when operating jointly within a multisensor fusion

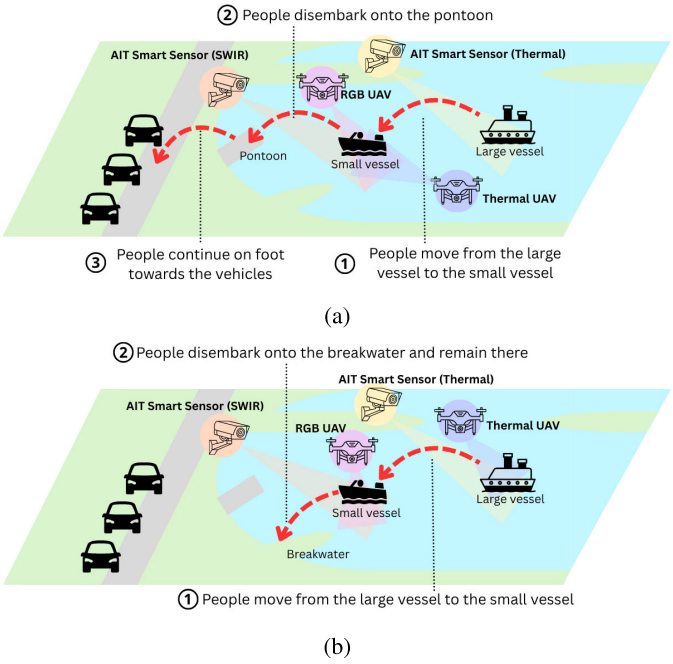


Fig. 11. Illustration of the scenarios conducted in the case study: (a) Scenario 1 and (b) Scenario 2.

system. The case study was performed in simulated scenarios within real-world environments. For each scenario, the AIT Smart Sensor, RGB UAV, and Thermal UAV platforms (GS-SWIR, GS-Therm, UAV-RGB, and UAV-Therm) were deployed to perform real-time surveillance tasks, including object detection, tracking, and geolocation approximation. The results produced by each platform were then fed into two systems: a nonfusion system and a fusion system. The nonfusion system used all detections from the individual sensor platforms without applying fusion, while the fusion system systematically combined detections across platforms using the multimodal data fusion method described in [50]. More details about this fusion method are provided below. The performance of the fusion system was then compared with the nonfusion system to assess whether combining multiple sensors offers advantages over using them separately.

1) Setup and Execution: Two scenarios were considered in the case study. They were designed to reflect maritime and coastal surveillance operations as follows.

1) *Scenario 1:* This scenario involves two vessels (one large and one small), multiple people, and several vehicles [see Fig. 11(a)]. The large vessel, transporting a group of people, rendezvouses with the small vessel, after which some individuals transfer to the small vessel (1). The small vessel subsequently returns to the shore, while the large vessel departs. Upon reaching a pontoon connected to the shore, the individuals on the small vessel disembark (2) and proceed toward nearby vehicles on the road (3).
2) *Scenario 2:* This scenario is similar to Scenario 1, except that the small vessel navigates toward and stops at a breakwater instead of the pontoon [see Fig. 11(b)]. The individuals disembark and remain on the breakwater (2).

TABLE IX
PARAMETERS OF FUSION FOR GRAPH G

Parameter	Value
UAV-RGB (τ, p)	(1, 0.5)
UAV-Therm (τ, p)	(1, 0.5)
GS-SWIR (τ, p)	(0.2, 0.3)
GS-Therm (τ, p)	(0.2, 0.3)
θ	0.8
Resolution (m)	5

During the scenario execution, GS-SWIR and GS-Therm were statically installed at different locations, as shown in Fig. 11. However, both sensors were able to pan, tilt, and zoom to track vessels. UAV-RGB and UAV-Therm operated dynamically, moving to track specific vessels or people depending on the scenario. For Scenario 1, GS-SWIR focused on the small vessel, while GS-Therm followed the large vessel. Both UAVs maintained focus on the small vessel. As for Scenario 2, GS-SWIR and UAV-Therm focused on the large vessel, while GS-Therm and UAV-RGB tracked the small vessel.

Seven GNSS trackers were used to collect ground-truth coordinates of the objects of interest for subsequent fusion evaluation. One tracker was placed on the large vessel, one on the small vessel, and five on the individuals who transferred from the large vessel to the small vessel and later disembarked on shore. Scenario 1 was repeated for seven trials, and Scenario 2 was repeated for five trials to ensure the reliability of the results.

2) Fusion Method: For the data fusion, we used a hierarchical fusion graph (HFG) $\mathcal{G}$ (see Fig. 12) as introduced in [51]. The information from each sensor is accumulated in sensor nodes, with the representation used being probabilistic occupancy maps (POMs). Particularly, a grid-based representation of the area of interest is used to accumulate sensor information by performing Bayesian filtering with a prior p and a temporal decay τ as first described in [50]. The POMs from individual sensors are then fused by performing a LogicalAND ($\wedge$) fusion, described in [52], between one UAV sensor (either UAV-RGB or UAV-Therm) and one ground-based sensor (either GS-SWIR or GS-Therm), yielding four different fused maps. These four maps are then fused with a Bayesian fusion [50] to produce one fused map (B). After that, we used thresholding with a threshold θ to define regions and image processing techniques to identify connected components, which serve as the fused detections. The final outputs from the fusion system are polygons of detected objects with classes and confidence scores, based on the information from multiple sensors. The parameters (see Table IX) and fusion graph (see Fig. 12) used are the same as in [52].

3) Evaluation Approach: We compared the performance of the nonfusion and fusion systems using the false positive rate (FPR), F1 Score, and the average center-to-center distance error between the estimated geolocations (either points or polygons) and the ground-truth coordinates. These metrics were calculated for detections produced by both systems.

4) Results: Table X presents the results of the nonfusion system compared with the fusion system. For Scenario 1, the

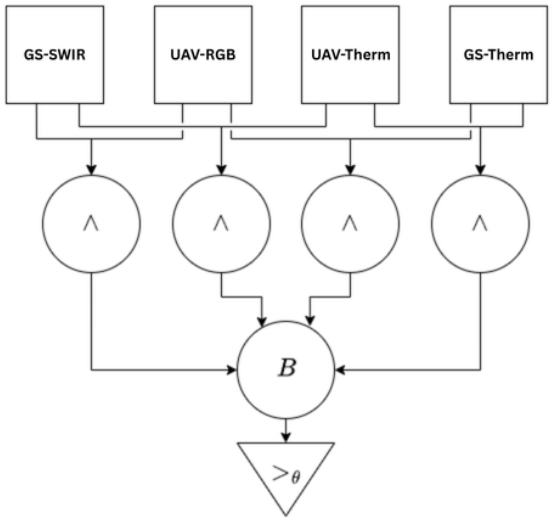


Fig. 12. Composition of fusion graph *G*.

TABLE X
FPR, F1-SCORE, AND ERROR DISTANCE (DIST.) FOR BOTH NONFUSION (N) AND FUSION (F) DETECTIONS ACROSS THE DIFFERENT TRIALS OF THE TWO SCENARIOS

Scenario	FPR (%)		F1 (%)		Dist. (m)	
	N	F	N	F	N	F
1.1	1.56	8.39	54.80	36.64	19.81	8.51
1.2	9.36	10.74	49.38	34.61	4.25	9.82
1.3	0.51	14.35	49.01	34.78	10.47	11.72
1.4	0.18	16.67	41.54	23.38	20.59	7.23
1.5	7.57	0.00	46.56	36.05	19.60	7.64
1.6	14.32	0.00	43.65	21.22	38.62	8.52
1.7	0.69	4.40	65.04	31.93	43.25	12.14
1	4.88	7.79	50.00	31.23	22.37	9.37
2.1	40.11	5.04	39.85	34.53	10.41	13.90
2.2	36.55	20.00	40.19	26.11	7.79	11.68
2.3	36.71	2.54	48.18	47.46	14.4	11.90
2.4	33.44	16.67	44.26	41.91	8.80	10.72
2.5	49.17	21.71	51.48	34.95	5.22	9.20
2	39.20	13.19	44.79	36.99	9.32	11.48

FPR of the fused detections was higher than that of the nonfusion detections, and the F1-score also decreased. This is due to the sensor executions not aligning with the fusion method configuration. In particular, as aforementioned, the fusion method requires detections from one UAV and one ground-based sensor to perform data fusion. However, in Scenario 1, only the ground-based sensor (GS-Therm) tracked the large vessel, while the other three sensors (GS-SWIR, UAV-RGB, and UAV-Therm) followed the small vessel. As a result, the fusion system did not produce improved results. However, in terms of distance error, the geolocations of fused detections were more accurate compared to the detections produced separately by individual sensor platforms. This indicates that combining estimations from multiple sensors reduces individual errors and leverages complementary information. This demonstrates the benefit of jointly using multiple sensors and the fusion method to refine the geolocation of detected objects.

For Scenario 2, the sensor execution was more closely aligned with the configuration of the fusion method; therefore, the improvement in detection accuracy from using multiple sensors can be observed. Using the fusion system, the FPR decreased from 39.20% to 13.19%. This corresponds to a 66.35% relative reduction of false alarms. The F1-Score, however, decreased from 44.79% to 36.99% (17.4% relative reduction), because some TPs were filtered out during fusion. Nonetheless, the reduction in FPR is much larger than the decrease in F1-Score. This demonstrates that multisensor fusion improves reliability by reducing false alarms with limited impact on true positive detections.

Moreover, the results in Table X show the difficulties of deploying sensor and fusion systems in real-world environments. In particular, the results of Scenarios 1 and 2 are largely different, despite using the same sensors and similar playbooks. The FPR was 4.88% in the first scenario but about eight times higher in the second, at 39.20%. The reason is that the trials were not performed in a controlled lab environment but on a coast, which can lead to other vessels being present in the area. We have observed that the ground-based sensors often detected vessels on the horizon, which can skew the results drastically. A vessel that is not part of the trial and that is correctly detected by the ground-based sensor will have no corresponding ground truth and therefore be classified as a false positive. The fusion methodology used tries to circumvent this problem by requiring detections from both ground-based and UAV platforms to classify a valid fused detection. This highlights both the importance as well as the difficulty in performing data fusion. A more detailed analysis of the difficulties in performing data fusion in maritime applications has been done in [52].

VIII. CONCLUSION

This work presents a comprehensive maritime surveillance system integrating four sensor platforms across multiple altitudes and sensing modalities: the AIT Smart Sensor platform (RGB, thermal, SWIR, and UV), the RGB UAV platform, the thermal UAV platform, and the satellite-based systems using Sentinel-1 (SAR) and Sentinel-2 (optical). Each was equipped with dedicated modules for object detection, tracking, and geolocation approximation, tailored to sensor-specific characteristics.

Performance was evaluated on newly acquired datasets across three tasks: object detection (covering persons, vessels, vehicles, and oil spills, where oil spill detection is exclusive to the RGB UAV platform), tracking, and geolocation approximation. In object detection, the AIT Smart Sensor platform consistently ranked among the top performers on ground-based imagery, particularly excelling in reducing FPs. The RGB UAV platform outperformed all baselines on aerial RGB imagery and showed strong results in oil spill detection. The thermal UAV platform performed well for detecting persons and vehicles, but showed some limitations in vessel detection compared to a state-of-the-art transformer-based detector; nonetheless, it remained effective across all classes. For the satellite-based systems, the SAR model demonstrated high precision and robustness under strict IoU thresholds, confirming its

suitability for maritime surveillance. In contrast, detection with optical satellite imagery proved more challenging due to resolution limits and environmental factors.

For object tracking, the AIT Smart Sensor platform demonstrated strong identity preservation and localization accuracy, performing better or comparably to state-of-the-art baselines, especially with challenging imagery data such as thermal and SWIR. The thermal UAV tracker showed competitive performance in balancing detection and association, though identity consistency remained weaker than state-of-the-art trackers such as BoT-SORT and ByteTrack coupled with a transformer-based detection model.

The geolocation algorithm for the thermal UAV platform achieved high accuracy, with errors under 11 m with optimal UAV camera pitch angles (lower than $-10°$) and under 45 m with near-horizontal pitch angles (higher than $-10°$). Incorporating horizon-aware projection and vertical scaling substantially reduced error, underscoring the importance of geometric corrections for reliable geolocation in thermal UAV imagery.

A case study was also conducted to examine the performance of the combined multisensor system with data fusion compared to individual sensor platforms. The results show that the multisensor system can leverage complementary information from different platforms, reducing false alarms and refining the geolocations of detected objects.

Overall, the results validate the effectiveness and highlight the unique advantages and limitations of each sensor platform. This work not only provides practical guidance for deploying multialtitude, multimodal sensor platforms in maritime environments but also lays the foundation for future research in multimodal data fusion and robust cross-platform surveillance systems.

Future work should expand scenario coverage to include broader maritime and land environments for richer annotations and explore multisensor fusion to capitalize on data from different platforms.

ACKNOWLEDGMENT

Thanet Markchom, Jonathan Boyle, Lulu Chen, and James Ferryman are with the Department of Computer Science, University of Reading, RG6 6UR Reading, U.K. (e-mail: thanet.markchom@reading.ac.uk; j.n.boyle@reading.ac.uk; l.chen@reading.ac.uk; j.m.ferryman@reading.ac.uk).

Olympia Kourounioti, George Voskopoulos, and Christos Kontopoulos are with Geosystems Hellas S.A., 11632 Athens, Greece (e-mail: o.kourounioti@geosystems-hellas.gr; g.voskopoulos@geosystems-hellas.gr; c.kontopoulos@geosystems-hellas.gr).

Matteo Marturini, Kilian Wohlleben, Stephan Veigl, Andreas Opitz, Dumitru Lunic, and Andreas Kriechbaum-Zabini are with Austrian Institute of Technology, 1210 Vienna, Austria (e-mail: Matteo.Marturini@ait.ac.at; Kilian.Wohlleben@ait.ac.at; Stephan.Veigl@ait.ac.at; Andreas.Opitz@ait.ac.at; Dumitru.Lunic@ait.ac.at; Andreas.Kriechbaum-Zabini@ait.ac.at).

Romaios Bratskas, Anastasios Gkamaris, and Dimitris Papachristos are with SKYLD Security and Defence Ltd., 2404 Nicosia, Cyprus (e-mail: rb@skyld.com.cy; a.gkamaris@skyld.com.cy; d.papachristos@skyld.com.cy).

George Leventakis is with the University of the Aegean, 83200 Samos, Greece (e-mail: george.leventakis@aegean.gr).

REFERENCES

[1] *Annual Risk Analysis 2025/2026*, Eur. Border Coast Guard Agency, Warsaw, Poland, 2025.

[2] J. N. Briggs, *Target Detection by Marine Radar*, vol. 16. Edison, NJ, USA: IET, 2004.

[3] C. Gamage, R. Dinalankara, J. Samarabandu, and A. Subasinghe, "A comprehensive survey on the applications of machine learning techniques on maritime surveillance to detect abnormal maritime vessel behaviors," *WMU J. Maritime Affairs*, vol. 22, no. 4, pp. 447–477, Dec. 2023.

[4] L. Patino, T. Cane, A. Vallee, and J. Ferryman, "PETS 2016: Dataset and challenge," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2016, pp. 1240–1247.

[5] J. Qu, R. W. Liu, Y. Guo, Y. Lu, J. Su, and P. Li, "Improving maritime traffic surveillance in inland waterways using the robust fusion of AIS and visual data," *Ocean Eng.*, vol. 275, May 2023, Art. no. 114198.

[6] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, "Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 8, pp. 1993–2016, Aug. 2017.

[7] Z. Shao, W. Wu, Z. Wang, W. Du, and C. Li, "SeaShips: A large-scale precisely annotated dataset for ship detection," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2593–2604, Oct. 2018.

[8] B. Carrillo-Perez, S. Barnes, and M. Stephan, "Ship segmentation and georeferencing from static oblique view images," *Sensors*, vol. 22, no. 7, p. 2713, Apr. 2022.

[9] N. Wang, Y. Wang, Y. Wei, B. Han, and Y. Feng, "Marine vessel detection dataset and benchmark for unmanned surface vehicles," *Appl. Ocean Res.*, vol. 142, Jan. 2024, Art. no. 103835.

[10] D. Qiao, G. Liu, T. Lv, W. Li, and J. Zhang, "Marine vision-based situational awareness using discriminative deep learning: A survey," *J. Mar. Sci. Eng.*, vol. 9, no. 4, p. 397, Apr. 2021.

[11] 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," *J. Mar. Sci. Eng.*, vol. 13, no. 5, p. 913, May 2025.

[12] J. Ding, W. Li, L. Pei, M. Yang, A. Tian, and B. Yuan, "Novel pipeline integrating cross-modality and motion model for nearshore multi-object tracking in optical video surveillance," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 9, pp. 12464–12476, Sep. 2024.

[13] F. Farahnakian and J. Heikkonen, "Deep learning based multi-modal fusion architectures for maritime vessel detection," *Remote Sens.*, vol. 12, no. 16, p. 2509, Aug. 2020.

[14] J. Jang, S. Oh, Y. Kim, D. Seo, Y. Choi, and H. J. Yang, "M^2sodai: Multi-modal maritime object detection dataset with RGB and hyperspectral image sensors," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2023, pp. 53831–53843.

[15] X. Bai, Z. Zhang, and X. Xu, "Advanced multispectral detection of maritime targets for unmanned ocean surveillance," *Ocean Eng.*, vol. 329, Jun. 2025, Art. no. 121185.

[16] W. Yong, L. Ling, T. Xiao, and J. Bing, "Multi-modal ship object detection in optical and SAR images," in *Proc. IEEE 2nd Int. Conf. Image Process. Comput. Appl. (ICIPCA)*, Jun. 2024, pp. 1321–1325.

[17] S. D. Priyadharshini and K. Vadivazhagan, "Enhanced vessel detection in maritime surveillance using multi-modal data integration and deep learning," in *Proc. 8th Int. Conf. I-SMAC (IoT Social, Mobile, Analytics Cloud) (I-SMAC)*, Oct. 2024, pp. 1090–1099.

[18] D. D. Bloisi, L. Iocchi, A. Pennisi, and L. Tombolini, "ARGOS-venice boat classification," in *Proc. 12th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2015, pp. 1–6.

[19] M. M. Zhang, J. Choi, K. Daniilidis, M. T. Wolf, and C. Kanan, "VAIS: A dataset for recognizing maritime imagery in the visible and infrared spectrums," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 10–16.

[20] L. Patino, T. Cane, and J. Ferryman, "A comprehensive maritime benchmark dataset for detection, tracking and threat recognition," in *Proc. 17th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Nov. 2021, pp. 1–8.

[21] R. Ribeiro, G. Cruz, J. Matos, and A. Bernardino, "A data set for airborne maritime surveillance environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 9, pp. 2720–2732, Sep. 2019.

[22] S. Yao et al., "WaterScenes: A multi-task 4D radar-camera fusion dataset and benchmarks for autonomous driving on water surfaces," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 11, pp. 16584–16598, Nov. 2024.

[23] G. Jocher and J. Qiu, "Ultralytics YOLO11," Ultralytics, Frederick, MD, USA, Tech. Rep., 2024.

[24] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "BoT-SORT: Robust associations multi-pedestrian tracking," 2022, *arXiv:2206.14651*.

[25] C. Liu, Y. Ding, H. Zhang, J. Xiu, and H. Kuang, "Improving target geolocation accuracy with multi-view aerial images in long-range oblique photography," *Drones*, vol. 8, no. 5, p. 177, Apr. 2024.

[26] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 21–37.

[27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[28] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *Int. J. Comput. Vis.*, vol. 56, no. 3, pp. 221–255, Feb. 2004.

[29] R. Torres et al., "GMES Sentinel-1 mission," *Remote Sens. Environ.*, vol. 120, pp. 9–24, May 2012.

[30] F. Spoto et al., "Overview of sentinel-2," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 1707–1710.

[31] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[32] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," 2024, *arXiv:2402.13616*.

[33] S. Yang, Z. Cao, N. Liu, Y. Sun, and Z. Wang, "Maritime electro-optical image object matching based on improved YOLOv9," *Electronics*, vol. 13, no. 14, p. 2774, Jul. 2024.

[34] Y. Zhang, Y. Yuan, Y. Feng, and X. Lu, "Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5535–5548, Aug. 2019.

[35] A. Agarwal. *Ship-Detection*. Accessed: Oct. 21, 2025. [Online]. Available: https://github.com/amanbasu/ship-detection/tree/master

[36] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, p. 765, Mar. 2019.

[37] T. Markchom et al., "PETS2025: Multi-authority multi-sensor maritime surveillance challenge and evaluation," in *Proc. IEEE Int. Conf. Adv. Vis. Signal-Based Syst. (AVSS)*, Aug. 2025, pp. 1–14.

[38] Copernicus Data Space Ecosystem. (2024). *Sentinel-2 Data*. Accessed: Apr. 30, 2025. [Online]. Available: https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data/sentinel-2

[39] G. Jocher, "Ultralytics YOLOV5," Ultralytics, Frederick, MD, USA, Tech. Rep., 2020.

[40] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

[41] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOV8," Ultralytics, Frederick, MD, USA, Tech. Rep., 2023.

[42] Y. Zhao et al., "DETRs beat YOLOs on real-time object detection," 2023, *arXiv:2304.08069*.

[43] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3645–3649.

[44] Y. Zhang et al., "ByteTrack: Multi-object tracking by associating every detection box," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2021, pp. 1–21.

[45] A. Milan, L. Leal-Taixe, I. Reid, S. Roth, and K. Schindler, "MOT16: A benchmark for multi-object tracking," 2016, *arXiv:1603.00831*.

[46] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 17–35.

[47] J. Luiten et al., "HOTA: A higher order metric for evaluating multi-object tracking," *Int. J. Comput. Vis.*, vol. 129, no. 2, pp. 548–578, Feb. 2021.

[48] E. Namazi, R. Mester, C. Lu, and J. Li, "Geolocation estimation of target vehicles using image processing and geometric computation," *Neurocomputing*, vol. 499, pp. 35–46, Aug. 2022.

[49] X. Chen, N. Su, Y. Huang, and J. Guan, "False-alarm-controllable radar detection for marine target based on multi features fusion via CNNs," *IEEE Sensors J.*, vol. 21, no. 7, pp. 9099–9111, Apr. 2021.

[50] M. Hubner et al., "A Bayesian approach—Data fusion for robust detection of vandalism and trespassing related events in the context of railway security," in *Proc. 27th Int. Conf. Inf. Fusion (FUSION)*, Jul. 2024, pp. 1–7.

[51] M. Hubner et al., "Robust detection of critical events in the context of railway security based on multimodal sensor data fusion," *Sensors*, vol. 24, no. 13, p. 4118, Jun. 2024.

[52] K. Wohlleben et al., "Enhancing maritime situational awareness through multimodal fusion: Insights from a real-world experiment," in *Proc. 17th Symp. Sensor Data Fusion, Trends, Solutions Appl.*, Nov. 2025, pp. 1–8.

Thanet Markchom received the Ph.D. degree in computer science from the University of Reading, Reading, U.K., in 2024.

He is currently a Postdoctoral Research Assistant with the Computational Vision Group, University of Reading. His research interests include computer vision, natural language processing, and recommender systems.

Olympia Kourounioti received the Diploma degree in rural, surveying, and geoinformatics engineering from the National Technical University of Athens, Athens, Greece, in 2018, and the joint M.Sc. degree in artificial intelligence and visual computing from the University of West Attica (UNIWA), Aigaleo, Greece, and the University of Limoges, Limoges, France, in 2023. She is currently pursuing the Ph.D. degree in remote sensing with applications in agriculture with the Geospatial Technology Laboratory, UNIWA.

She has been with Geosystems Hellas S.A., Athens, Greece, since 2023, participating in various EU Horizon Projects as a Remote Sensing and Geoinformatics Engineer and a Project Manager. Her research interests include advanced processing, classification, and fusion of multispectral and hyperspectral remote sensing data using machine learning and deep learning methodologies.

Matteo Marturini has been a Research Engineer with the Data Science and Artificial Intelligence Group, Austrian Institute of Technology, Austria, since April 2024. He has developed solid expertise in computer vision and contributes actively to European projects on computer vision and scene understanding, with a focus on object detection, multiobject tracking, and vision-language models for real-world applications.

Romaios Bratskas received the B.Sc. degree from the University of Macedonia, Thessaloniki, Greece, in 2005, and the M.B.A. degrees in economics from European University Cyprus, Nicosia, Cyprus, in 2008.

He is the Chief Operating Officer at SKYLD Security and Defence Ltd., Cyprus. He leads business operations and strategic development, drawing on experience as a project manager, business analyst, and consultant. His expertise includes strategic planning and marketing, communication strategy and public affairs, and the management of EU and national research and development programmes and multipartner projects.

Kilian Wohlleben studied Financial Mathematics at the LMU Munich, Munich, Germany, and the University College London, London, U.K.

Afterwards, he worked in the insurance industry for two years before joining Austrian Institute of Technology, Austria, as a Research Engineer in 2022. There, he conducts research into algorithms for distributed acoustic sensing and data fusion.

Jonathan Boyle photograph and biography not available at the time of publication.

Lulu Chen received the Ph.D. degree in computer science from the University of Reading, Reading, U.K., in 2012.

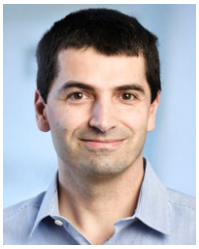
She has been contributing to several European research and innovation projects. Her research interests include computer vision, biometric recognition, behaviour analysis, and risk assessment in security and surveillance applications.

George Voskopoulos received the master's in Informatics and Computer Engineering degree in informatics and computer engineering from the University of West Attica, Aigaleo, Greece, in 2023.

He has been a Geospatial and a Full-Stack Engineer with Geosystems Hellas S.A., Athens, Greece, since 2022. His expertise spans spatial data analysis, system integration, and end-to-end application development.

Christos Kontopoulos received the M.Sc. degree in rural, surveying, and geoinformatics engineering from the National Technical University of Athens (NTUA), Athens, Greece, in 2015, specializing in Earth observation, remote sensing, AI and machine learning, geographic information systems (GIS), photogrammetry, and geospatial data analysis.

He is the Chief Technology Officer at Geosystems Hellas S.A. (GSH), Athens. With extensive experience in research and technology development, he has participated in numerous projects funded by EC and ESA, leading technical teams and overseeing system implementations. His work spans the design, development, and support of GIS, photogrammetry, and remote sensing systems, focusing on the effective use of geospatial data for technical, research, and operational applications. He is leading the company's engineering team, driving innovation and advancing the adoption of cutting-edge geospatial technologies.



Stephan Veigl received the M.Sc. degree in computer science from Vienna University of Technology (TU Wien), Vienna, Austria.

He is a Project Manager and a Research Engineer at Austrian Institute of Technology (AIT), Austria. He has co-authored several peer-reviewed journal and conference publications and has project experience in border surveillance, counter-terrorism, and critical infrastructure protection. His research interests include video processing system architectures, computer vision, multimodal sensor data fusion, and machine learning, with a focus on multimodal sensor processing and the robust detection of critical events in complex environments for surveillance and security applications.

Andreas Opitz received the M.Sc. degree in computer science from the Vienna University of Technology, Wien, Austria, in 2009.

He is working as a Research Engineer at the AIT Austrian Institute of Technology, Vienna, Austria, in the field of applied research in computer vision and machine learning. His work focuses on the design and development of sensor systems for safety and security applications. He has co-authored several journal and conference publications addressing challenges in visual sensing under real-world conditions.

Anastasios Gkamaris received the Bachelor of Applied Science degree in mechatronics, robotics, and automation engineering from the Technological Educational Institute of Piraeus, Athens, Greece, in 2011.

He holds an EASA A1-A2-A3-STS01-02 UAV Pilot License. He is currently involved in projects in the security, defense, and critical infrastructure inspection sectors. With over 15 years of experience in field engineering, he specializes in systems integration, UAV operations, and technical project execution. His research interests include uncrewed aerial systems and surveillance applications.

Dimitris Papachristos received the M.Sc. degree in data communication (Hons.) from Kingston University (UK) – TEI Piraeus (Greece), in 2005, specialization in home networking, interdepartmental postgraduate programme. He is pursuing the the Ph.D. (Doctorate) degree with the Department of Shipping, Trade and Transport, University of the Aegean, Samos, Greece, in June 2018, and the Ph.D. degree candidate with Department of Port Management and Shipping, National and Kapodistrian University of Athens (NKUA), Athens, Greece.

He is a M.Sc. Programme (Bioethics) at Medical School, Democritus University of Thrace (DUTH), Alexandroupoli, Greece, the M.Sc. Programme (Advanced Systems and Methods in Biomedical Technology) at the Department of Biomedical Engineering, University of West Attica, Athens, Greece. From 2010to 2013, he was a M.Sc. Programme ("Technoglossia" / Language Technology) at the National and Kapodistrian University of Athens (NKUA), National Technical University of Athens (NTUA), Athens. From 2005 to 2009, he was a M.Sc. Programme (ICT in Education / Communication and Information Technologies for Education) at the National and Kapodistrian University of Athens (NKUA) – University of Thessaly – Piraeus University of Applied Sciences (TEI Piraeus), Greece, from September 2020 to Present, he is a Postdoctoral Researcher with the Department of Shipping, Trade and Transport, University of the Aegean, Samos. He is a Consultant-External Associate at SKYLD Security and Defence Ltd., Cyprus. He has contributed to EU-funded research and development projects spanning defence, security, and industrial applications. His interests include education technology, automation, data analysis, and digital transformation.

Dumitru Lunic is pursuing the Bachelor of Science degree in informatics with the IMC FH Krems, Krems an der Donau, Austria.

He is currently an Intern at Austrian Institute of Technology, Austria, where he focuses on data fusion and data visualization. His research interests include machine learning and practical applications of data science.

James Ferryman (Member, IEEE), photograph and biography not available at the time of publication.

Andreas Kriechbaum-Zabini is a Thematic Coordinator for the research field "computer vision" at Austrian Institute of Technology (AIT), Austria. His research focuses on advanced sensor systems, AI-based situational awareness, and data fusion with a focus on border security and public safety and security applications. He has extensive experience in coordinating and contributing to European research and innovation projects. His work addresses operational surveillance, interoperability, and decision-support systems for security authorities. He collaborates closely with industry, academia, and governmental stakeholders.

George Leventakis received the M.Sc. degree in risk management from the University of Technology Sydney, Sydney, Australia, in 2000, the M.B.A. degree from Athens University of Economics and Business, Athens, Greece, in 2003, and the Ph.D. degree in operational risk assessment modeling from the University of the Aegean, Samos, Greece, in 2013.

He is a Senior Advisor in EU R&D at SKYLD Security and Defence Ltd., Cyprus, with long-standing experience in security technology, critical infrastructure protection, and project governance. His work includes proposal development, consortium building, and supporting the exploitation of research results into operational solutions.