AIRSHIP

Funded by the European Union's

Horizon Europe programme under



D6.2 Report on Real-time multi-modal sensor fusion for awareness in WIGs and its applications for fast sense-and-avoid perception and Real time sea wave perception and modelling

WP6 Task

Disclaimer

The content of this deliverable reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

Document Information

Grant Agreement / Proposal ID	101096487
Project Title	AUTONOMOUS FLYING SHIPS FOR INTER-ISLAND AND INLAND WATERS TRANSPORT
Project Acronym	AIRSHIP
Project Coordinator	Universidad Politécnica de Madrid
Project starting date (duration)	1st January 2023 (48 months)
Related Work Package	Work Package 6 - Perception avionics
Related Task(s)	Task 6.2 Real-time multi-modal sensor fusion for WIGs
Lead Organisation	INESCTEC
Contributing Partner(s)	Universidade Politécnica de Madrid (UPM)
Due Date	31st December 2025
Submission Date	31st December 2025
Dissemination level	PU

Revision History

Revision	Revision Date	Name	Position
1	15/12/2025	José M. Almeida	Researcher
2	16/12/2025	André Dias	Researcher
3	17/12/2025	Pedro Lima	Researcher

TABLE OF CONTENTS

1 Introduction 8
2 AIRSHIP Perception Elements 8
3 Localisation system 9
4 Multi-modal fusion 11
5 Multi-modal sensor fusion 12
5.1 Update step with JPDA and adaptive measurement noise 15
5.2 Data association 16
5.3 Track Management 17
6 Foundational Sensor Processing 18
6.1 Registration and Synchronization 18
6.2 Sensor Calibration 19
6.3 Pre-processing 20
6.3.1 Range Sensor Clustering 20
6.3.2 Visible Image Processing 22
6.3.3 Thermal Image Processing 23
7 Tests and Results 25
7.1 Field tests 25
7.1.1 Field Test Results 26
7.2 Simulation 30
7.2.1 Simulation Scenario 30
7.2.2 Sensor-level target detections 32
7.2.3 Simulation Results 35
8 Surface Modelling 41
8.1 Wave Detection Technologies 42
8.2 Linear Wave Theory 42
8.3 WIG Sensors for Wave Perception 43
8.4 Current Results 44
9 Conclusion 46
References 47

Funded by the
European Union

List of Tables

Table 1. Abbreviations and Acronyms..... 6
Table 2 - Perception system elements..... 9
Table 3 - Sensor synchronization method..... 19
Table 4 - 2D Euclidean Error Statistics..... 28
Table 5 - 3D Euclidean Error Statistics..... 28
Table 6 - Comparison between simulated and real sensors..... 32

List of Figures

Figure 1 - Localization system architecture..... 10
Figure 2 - Multi-modal approaches..... 11
Figure 3 - Perception system architecture..... 12
Figure 4 - Diagram of the implemented EKF state estimator..... 13
Figure 5 - Flowchart of the Joint Probabilistic Data Association approach..... 16
Figure 6 - Track management module flowchart..... 17
Figure 7 - Synchronization architecture..... 18
Figure 8 - Original and undistorted camera image..... 19
Figure 9 - Stereo pair calibration..... 20
Figure 10 - LiDAR clustering output..... 21
Figure 11 - RADAR clustering output..... 21
Figure 12 - Target detection on the input image..... 22
Figure 13 - Point cloud reconstruction from a stereo pair..... 22
Figure 14 – 3D LiDAR point cloud projected on the 2D thermographic image..... 24
Figure 15 - Thermal image processing architecture..... 24
Figure 16 - PORTUS USV..... 25
Figure 17 - Perception system's diagram used in field missions..... 26
Figure 18 - Multi-modal fusion diagram..... 26
Figure 19 - Different Kalman configurations trajectories using CV model compared to ground-truth..... 27
Figure 20 - Different Kalman configurations trajectories using CA model compared to ground-truth..... 27
Figure 21 - 2D Euclidean error of Kalman output using LiDAR-only, RADAR-only, and fused measurements for CA and CV motion models..... 29
Figure 22 - 3D Euclidean error of Kalman output using LiDAR-only, RADAR-only, and fused measurements for CA and CV motion models..... 29
Figure 23 - WIG model used in the simulation and its sensors..... 30
Figure 24 - Obstacles used in simulations..... 31
Figure 25 - WIG model trajectory in the simulation. On the left, the loitering trajectory, and on the right, the straight trajectory..... 31
Figure 26 - Point cloud generated by stereo camera pair when the WIG is static..... 32
Figure 27 - Point cloud generated by stereo camera pair when WIG is in motion..... 33
Figure 28 - Example of clustered point cloud showing obstacle detections. Different colours represent distinct targets..... 33

Figure 29 - Camera image (left) and raw LiDAR detections (right). 34
Figure 30 - Camera image (left) and LiDAR detections (right). Different colours represent distinct targets. 34
Figure 31 - Camera image (left) and raw RADAR detections (right). 35
Figure 32 - Camera image (left) and RADAR detections after Euclidean clustering (right). Different colours represent distinct targets. 35
Figure 33 - Depth offset between the obstacle centre and the possible detection centroid. 36
Figure 34 - Visual representation of the centroid discrepancy depending on the viewing angle of the target ... 36
Figure 35 - Ground truth trajectories of the four vessel targets and WiG in NED coordinates 37
Figure 36 - Ground truth trajectories of the four vessel targets – target one (upper left), target two (upper right), target three (lower left) and target four (lower right) – vs. the Kalman filter output for straight path trajectory in NED coordinates. 38
Figure 37 - Distance errors between ground truth and Kalman filter estimates for each vessel during the straight path trajectory 39
Figure 38 - Ground truth trajectories of the four vessel targets – target one (upper left), target two (upper right), target three (lower left) and target four (lower right) – vs. the Kalman filter output for orbital trajectory in NED coordinates. 40
Figure 39 - Distance errors between ground truth and Kalman filter estimates for each vessel during the orbital trajectory 41
Figure 40 - Sea wave reconstruction from stereo vision. 44

Abbreviations and Acronyms

Acronym	Description
AIS	Automatic Identification System
CA	Constant Acceleration
CNN	Convolutional Neural Network
CV	Constant Velocity
EKF	Extended Kalman Filter
GNN	Global Nearest Neighbour
GNSS	Global Navigation Satellite System
IMU	Inertial Measurement Unit
JPDA	Joint Probabilistic Data Association
LiDAR	Light Detection and Ranging
NED	North East Down

PPS	Pulse Per Second
PTP	Precision Time Protocol
RADAR	Radio Detection and Ranging
RTK	Real-Time Kinematic
SWaP	Size Weight and Power
USV	Unmanned Surface Vehicle
WIG	Wing-in-Ground
WP	Work Package

Table 1. Abbreviations and Acronyms

EXECUTIVE SUMMARY

This report provides an overview of the progress made within Work Package (WP) 6 of the AIRSHIP project. It focuses on the research and development of advanced sensing and perception methods for autonomous WIG drones.

The primary objective of WP6 is to equip the AIRSHIP systems with advanced perception and awareness capabilities. This goal, based on the deliverable D6.1, will promote low-latency and long-range detection and tracking of obstacles, thereby sustaining the effective implementation of collision avoidance strategies. Furthermore, WP6 investigates multi-sensor approaches for accurate and real-time perception of sea waves and environmental modelling to ensure optimal flight conditions.

The primary focus of this deliverable is Task 6.2, Real-time Multimodal Sensor Fusion for WIGs, which is concluded with this report. Task 6.2 involved the study, development, and testing of real-time sensor fusion approaches that combine Radio Detection and Ranging (RADAR), Light Detection and Ranging (LiDAR), optical and thermal cameras, and high-rate navigation sensors to build a detailed, dynamic model of the environment. The task successfully met its objective of providing low-latency and long-range obstacle detection and tracking capabilities, forming the perception basis required for the collision-avoidance mechanisms developed in WP7.

Additionally, this document reports the ongoing progress of Task 6.3, which continues until Month 42 (M42). Task 6.3 focuses on studying and validating multi-sensory approaches for accurate and real-time floor perception and modelling (e.g., sea waves and orography), aimed at supporting optimal and safe WIG flight conditions. Although still in progress, preliminary developments and intermediate results are presented in this report.

1 Introduction

Europe's many archipelagos and inland waterways present mobility challenges that conventional transport modes struggle to address. Aircraft provide speed, but at a high economic and environmental cost, while ships offer capacity, albeit with limited speed and efficiency.

The AIRSHIP project addresses this gap by developing Wing-in-Ground (WIG) effect vehicles, which are flying ships that operate close to the surface, benefiting from increased lift and reduced drag. These vehicles aim to combine airborne speed with greater energy efficiency and reduced environmental impact, forming the basis for a new class of fully electric Unmanned WIG Vehicles.

Within this context, Work Package 6 develops the sensing and perception capabilities required for safe autonomous WIG operation. Its objectives include enabling low-latency, long-range obstacle detection and accurate real-time modelling of the surrounding environment, both of which are essential for collision avoidance and stable flight in ground effect.

This deliverable reports the progress achieved across the WP6 tasks to date. Task 6.1, completed and delivered in D6.1, established the sensor acquisition and synchronization framework. Task 6.2, which concludes with this document, focused on real-time multi-modal sensor fusion using radar, LiDAR, optical, and thermal cameras, as well as navigation sensors to create a unified environmental perception model. Task 6.3, ongoing until Month 42, investigates multi-sensor approaches for real-time floor and sea-surface perception to support optimal ground-effect flight.

The document is structured as follows: Section 2 summarizes the perception element; Section 3 describes the architecture and operation of the localization system; Section 4 presents the multimodal sensor fusion framework, including filtering, state update, data association, and target management; Section 5 details the foundational aspects of sensor processing, such as calibration procedures and image and range sensor data processing; Section 6 reports on perception system testing, covering both field experiments and simulation-based evaluations; Section 7 discusses surface modelling approaches; and Section 8 concludes the document and outlines final remarks.

2 AIRSHIP Perception Elements

Following the sensor selection presented in Deliverable D6.1, several components were modified or added. These changes are driven by analysis of the results achieved thus far, highlighting their necessity for future success.

Regarding LiDAR, the chosen unit, based on the quality of the generated point cloud, was the LsLidar LS180S2, replacing the previously considered Livox HAP T1/Tx, which offered a shorter range and a lower point rate. This choice aims to provide a higher-detail output and, consequently, improved performance in obstacle detection.

The Automatic Identification System (AIS) is a maritime communication technology designed to enhance situational awareness and navigational safety. By continuously broadcasting a vessel's identity, position, speed, and course, AIS enables ships and coastal stations to monitor surrounding traffic in real time. This information, provided every 2 to 10 seconds (3 minutes for slower vessels), plays a crucial role in collision avoidance, traffic management, and supporting autonomous or semi-autonomous maritime operations. This element was not included in Deliverable D6.1 because no suitably compact solution had been identified at that stage. In the meantime, the dAISy Mini AIS receiver was selected, as it not only meets the size constraints but also provides a low-power solution appropriate for the system.

The Global Navigation Satellite System (GNSS) receiver was also changed in response to recent interference observed in the operating environment (likely related to the ongoing conflict between Ukraine and Russia). The adopted solution differs from the original one in several key aspects, including support for dual-antenna configurations, integration with a wider range of satellite constellations, direct heading estimation capabilities, and enhanced positioning accuracy. This hardware permits only trusted satellites, such as Galileo (an EU-owned system). Combined with data from the Inertial Measurement Unit (IMU), this configuration enables precise state estimation at a high update rate.

Regarding the IMU, it was changed to a unit identified as having more appropriate characteristics than the one previously chosen. The SBG's Pulse-40 and the GuideNav's GUIDE730 are both tactical-grade MEMS IMUs, but they emphasize slightly different strengths. Pulse-40 focuses on Size Weight and Power (SWaP) and dynamic performance. It's extremely compact, offering high bandwidth (up to 500 Hz) and data rate (2 kHz), as well as ultra-low noise gyros and excellent vibration rectification performance, making it ideal for tight spaces and highly dynamic, vibrating platforms. In contrast, the GUIDE730 targets rugged precision over long periods, with tactical-grade gyro bias instability around 0.2°/h, a wide ±450°/s range, advanced temperature/misalignment/non-linearity compensation, attitude outputs, and robust operation from −40°C to +85°C under shock, vibration, and harsh electromagnetic conditions, all via a simple RS422 interface. Given these trade-offs, GUIDE730 is the more suitable choice when long-term bias stability, built-in error compensation, and straightforward system integration are more critical than minimal size alone, especially for navigation-driven robotics or autonomous platforms operating in demanding environments.

Table 2 - Perception system elements.

Component	Brand and Model	Description
LiDAR	LsLidar LS180S2	Range of 250 m with an accuracy of ± 2 cm. Generates 1.3 million points a second.
RADAR	ARS 548 RDI	Up to 1500 meters range. Wide FOV (±60° azimuth, ±14° elevation)
EO Camera	Dalsa Genie Nano GigE	3.1MP. Supports frame rates up to 53 Hz with TurboDrive.
IR Camera	Teledyne Calibir GXM	73.2° HFOV and a 54.4° VFOV. Detects temperatures from -45°C to 65°C.
Altimeter	Benewake TF02-i	0.1 to 13.5 m @ 10% reflectivity. Supports frame rates up to 100 Hz.
AIS Receiver	dAISy Mini	Low dimensions and power consumption. Dual channel AIS, NMEA ouput, VHF reception
IMU	GuideNav GUIDE730	Low gyro noise (0.03°/√h) and bias instability (0.2°/h) with high-precision accelerometers (10µg)
GNSS Receiver	ArduSimple simpleRTK3B Pro	Millimetre-level precision with high frame rate (100 Hz)

3 Localisation system

Given the prevalence of deliberate interference affecting GNSS systems, relying directly on unprocessed GNSS data is increasingly inadvisable. Furthermore, for mobile platforms, GNSS alone cannot provide the level of positional accuracy required. To address these limitations, GNSS measurements are combined with other sensor inputs. For the AIRSHIP, as shown in Figure 1, the sensor suite includes an IMU, a magnetometer and a barometer. During the initialization stage, conducted while the vehicle remains stationary, the system utilizes GNSS data, IMU readings, and heading information supplied by the magnetometer. After initialization, an Extended Kalman Filter (EKF) is employed to estimate the vehicle's position, velocity, and orientation. The EKF operates through two main steps: prediction and update. The prediction relies exclusively on IMU measurements, whereas the update phase integrates observations from the GNSS, barometer, and magnetometer. Although these sensors provide measurements at different frequencies, this does not introduce inconsistencies, as the EKF inherently accounts for each measurement's timestamp.

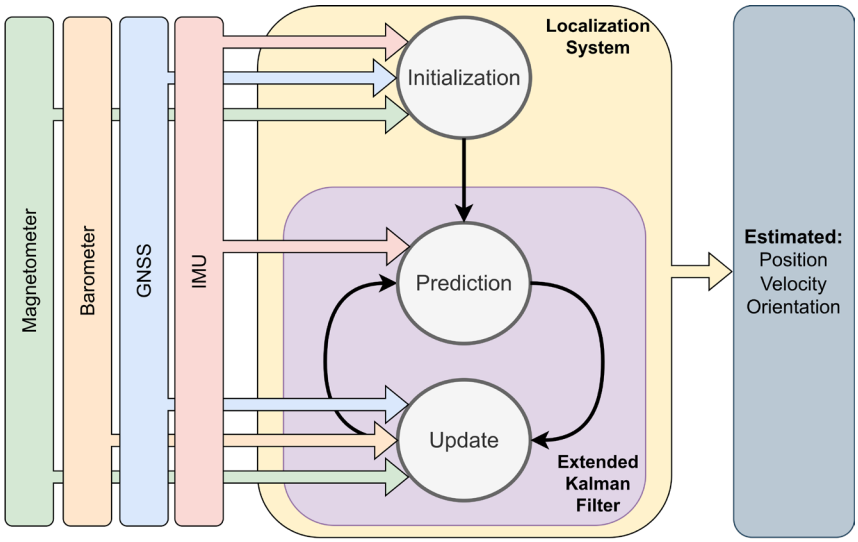


Figure 1 - Localization system architecture.

4 Multi-modal fusion

Individually, sensors cannot guarantee reliable and precise perception in all operating conditions. By combining complementary sensors in a multi-modal fusion approach, the system leverages redundancy to increase robustness, improve accuracy through cross-validation, and create a more comprehensive understanding of the system (in the case of fusion of navigation sensors) or the surrounding environment (Hu, et al., 2020) (Cho, Seo, Kumar, & Rajkumar, 2014) (Rodríguez Flórez, Frémont, Bonnifait, & Cherfaoui, 2014) (Helgesen, Brekke, Helgesen, & Engelhardtsen, 2019).

Multi-modal sensor fusion can occur at different stages of the processing pipeline, as illustrated in Figure 2 (Cooper, Zhang, Ragai, & Gao, 2024).

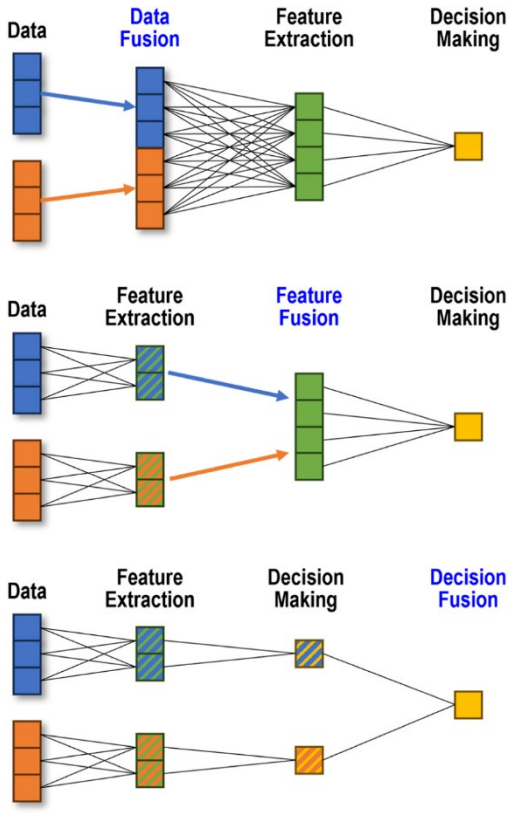


Figure 2 - Multi-modal approaches.

According to the literature, multi-modal sensor fusion can be categorized into three levels:

- Early fusion (sensor or measurement level) – Combines raw data from multiple sensor sources before processing. For this reason, this method offers rich information, but it is typically noisier, requiring robust calibration and synchronization. Because it uses raw data, without any filtration or segmentation, it can be computationally expensive.
- Intermediate fusion (feature level) – Combines features extracted from each sensor. Creates a balance between performance and complexity, which requires robust calibration and synchronization.
- Late fusion (decision level) – Refers to the combination of decisions (e.g., object classification) of each sensor. It is typically the simpler and robust fusion method; however, it also loses a large part of the sensor data since it was already pre-processed, segmented, or filtered.

In practice, many systems adopt hybrid fusion strategies that combine different levels depending on the task and computational constraints. Multi-modal systems introduce challenges such as managing heterogeneous data types, calibrating and synchronizing sensors with different data rates, and balancing the computational cost of real-time operation. However, only the integration of complementary modalities can provide the robustness and accuracy required for multi-target tracking.

5 Multi-modal sensor fusion

The continuous advancement of technology has enabled the development of increasingly smaller sensors with ever more impressive capabilities. Among these are imaging sensors, which over time have allowed the capture of images at progressively higher resolutions; LiDAR systems, which generate highly detailed point clouds capable of identifying features such as road lanes or pavement markings; and RADAR systems, which can now achieve longer ranges while operating at lower power levels. However, no single sensor can be considered "the best." Depending on the scenario, some sensors will outperform others.

The success of any perception system relies on the use of heterogeneous sensors so that, in every situation, at least one of them is able to accurately perceive the environment. Multi-modal sensor fusion enriches the overall information by leveraging the strengths and complementary contributions of each sensor type.

In the AIRSHIP project, developing a perception system is particularly challenging because the obstacles are unknown, weather conditions may be adverse, and the requirement for the platform to fly at low altitude creates an unusual operating scenario. Taking this into account, the sensor suite, as introduced in Section 2, includes electro-optical cameras (a pair, enabling stereo vision), infrared cameras, LiDAR systems, and RADAR.

After each sensor publishes its list of detected obstacles, some of which require dedicated processing blocks to extract obstacles and their corresponding positions from raw data, a data association module is executed. This last, which also counts on the system pose and information provided by the AIS, interacts with the track management component and an Extended Kalman Filter, which estimates the state of the obstacle(s). This architecture is presented in Figure 3.

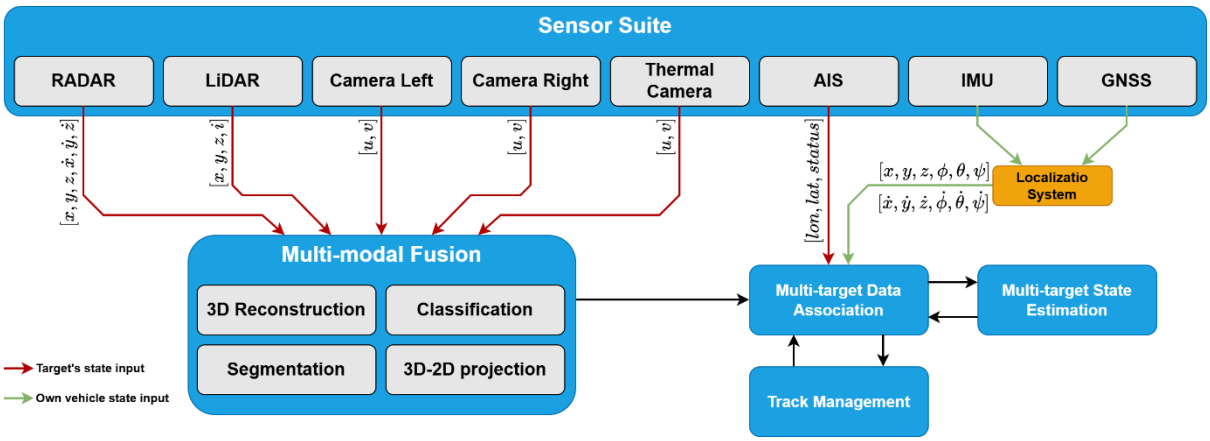


Figure 3 - Perception system architecture.

The EKF was selected as the state estimator because the motion model is linear, while the framework accommodates potential nonlinearities in future observation models with additive Gaussian noise. Under linear conditions, the EKF reduces to the optimal minimum mean-square-error estimator. It further provides linearization for nonlinear cases. EKF provides more flexibility in the chosen motion model, which can be subject to change in the future due to implementation in the WIG or depending on the target tracking scenarios.

An advantage of the EKF in this context, as with other Kalman-based state estimators, is its ability to maintain an explicit estimate of the full target state vector and its associated covariance matrix. The state variables are a subset of the system variables that can represent the entire state of the system at any given time (e.g., position, velocity, acceleration), while the covariance quantifies the estimator's uncertainty. Alternative methods, such as Random Finite Sets (RFS) approaches, treat the multi-target state as set-valued random variables. These methods propagate the intensity or cardinality distribution of the entire object set and are particularly useful in highly cluttered scenarios with a varying number of targets. However, they model the group of targets as a whole, rather than maintaining an explicit state estimate with covariance for each individual track.

The multi-target tracking problem is formulated as a nonlinear state estimation task, where the objective is to infer the continuous-time kinematic states of multiple dynamic obstacles from noisy and potentially ambiguous sensor measurements. Each target is modelled independently at the state level, while uncertainty from measurement ambiguity is handled at the update stage through the Joint Probabilistic Data Association (JPDA) method. The filter operates at a fixed rate of 20 Hz, ensuring consistent and timely state estimates for dynamic obstacle tracking. A high-level flowchart of the implemented EKF is shown in Figure 4.

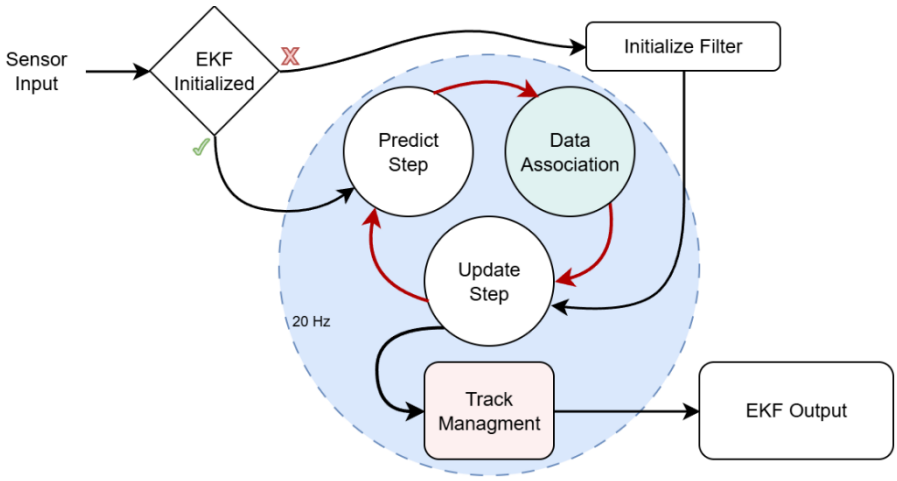


Figure 4 - Diagram of the implemented EKF state estimator.

The EKF employs two motion models to describe target dynamics: the Constant Velocity (CV) and the Constant Acceleration (CA) models. These provide a trade-off between modelling simplicity and motion fidelity, enabling the tracker to adapt to different target behaviours.

In a CV model, the target is assumed to move with constant velocity during the prediction interval Δt. The state vector is defined as:

$$\mathbf{x}_{CV} = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^T$$

Where (x, y, z) represent the target's position in Cartesian coordinates within the local fixed NED frame, and $(\dot{x}, \dot{y}, \dot{z})$ the corresponding velocity components. This model is suitable for targets that perform nearly uniform motion between updates.

The state transition model, which describes the evolution of the target state between consecutive time steps, is expressed in linear form as:

$$\mathbf{x}_{k|k-1} = F \cdot \mathbf{x}_{k-1|k-1} + w_k$$

Where F is the state transition matrix and w_k is the zero-mean Gaussian process noise. Given the CV state vector, the kinematic equations that describe this model can be formulated as:

$$\begin{cases} x_{k+1} = x_k + \dot{x}_k \Delta t \\ y_{k+1} = y_k + \dot{y}_k \Delta t \\ z_{k+1} = z_k + \dot{z}_k \Delta t \\ \dot{x}_{k+1} = \dot{x}_k \\ \dot{y}_{k+1} = \dot{y}_k \\ \dot{z}_{k+1} = \dot{z}_k \end{cases}$$

These equations can be compactly represented using the state transition matrix:

$$F_{CV} = \begin{bmatrix} I_3 & \Delta t \cdot I_3 \\ 0_3 & I_3 \end{bmatrix}$$

Where I_3 is the three-dimensional identity matrix and 0_3 is the three-dimensional zero matrix.

On the other hand, the CA model introduces an additional level of motion dynamics by incorporating acceleration terms into the state vector:

$$\mathbf{x}_{CA} = [x, y, z, \dot{x}, \dot{y}, \dot{z}, \ddot{x}, \ddot{y}, \ddot{z}]^T$$

The inclusion of acceleration components $(\ddot{x}, \ddot{y}, \ddot{z})$ allows the model to capture changes in velocity, making it more appropriate for manoeuvring targets or non-uniform motion.

The discrete-time kinematic equations for the CA model are given by:

$$\begin{cases} x_{k+1} = x_k + \dot{x}_k \Delta t + \frac{1}{2} \ddot{x}_k \Delta t^2 \\ y_{k+1} = y_k + \dot{y}_k \Delta t + \frac{1}{2} \ddot{y}_k \Delta t^2 \\ z_{k+1} = z_k + \dot{z}_k \Delta t + \frac{1}{2} \ddot{z}_k \Delta t^2 \\ \dot{x}_{k+1} = \dot{x}_k + \ddot{x}_k \Delta t \\ \dot{y}_{k+1} = \dot{y}_k + \ddot{y}_k \Delta t \\ \dot{z}_{k+1} = \dot{z}_k + \ddot{z}_k \Delta t \\ \ddot{x}_{k+1} = \ddot{x}_k \\ \ddot{y}_{k+1} = \ddot{y}_k \\ \ddot{z}_{k+1} = \ddot{z}_k \end{cases}$$

This model can be written in matrix form as:

$$F_{CA} = \begin{bmatrix} I_3 & \Delta t \cdot I_3 & \frac{1}{2} \Delta t^2 \cdot I_3 \\ 0_3 & I_3 & \Delta t \cdot I_3 \\ 0_3 & 0_3 & I_3 \end{bmatrix}$$

The CV model offers a computational efficiency choice for nearly uniform motion, while CA provides improved tracking for manoeuvring targets by explicitly modelling acceleration. Both formulations maintain linearity, allowing EKF to incorporate nonlinear observation functions through local linearization when required.

For each target, the predicted state estimate and covariance at time k are computed as:

$$\hat{\mathbf{x}}_{k|k-1} = F \cdot \hat{\mathbf{x}}_{k-1|k-1}$$

$$P_{k|k-1} = F \cdot P_{k-1|k-1} \cdot F^T + Q$$

Where $\hat{x}_{k|k-1}$ is the predicted state vector, $P_{k|k-1}$ is the predicted covariance matrix, F is the state transition matrix (dependent on the chosen motion model) and Q is the processed noise covariance, which represents the uncertainties associated with the model.

The observation model defines how the true target state is mapped to the measurements obtained from each sensor. In the current implementation, two types of sensors are considered: one providing three-dimensional position (x, y, z), and another providing both position and velocity measurements in the x and y axis $(x, y, z, \dot{x}, \dot{y})$. Each observation model is linear and can be expressed as:

$$\mathbf{z}_k = h(\mathbf{x}_k) + \mathbf{v}_k$$

Where $\mathbf{z}$ is the $h(\cdot)$ maps the target state to the measurement space and $\mathbf{v}_i$ is zero-mean Gaussian measurement noise. Because the function $h(\cdot)$ is nonlinear the EKF uses Jacobians evaluated at the current estimate:

$$\mathbf{H}_k = \frac{\partial h}{\partial \mathbf{x}} |_{\hat{x}_{k|k-1}}$$

Applying this equation to the two types of sensors previously mentioned, the following observation matrices are obtained:

$$\mathbf{H}_{pos} = [I_3, 0_{3x6}]$$

$$\mathbf{H}_{pos+vel} = [I_5, 0_{5x4}]$$

Both observation models remain linear and can therefore be directly integrated into the EKF measurement update equations. Any future introduction of nonlinear sensor measurements can be incorporated by extending these linear forms through first-order linearization around the current state estimate.

5.1 Update step with JPDA and adaptive measurement noise

The EKF implementation is extended to incorporate the JPDA framework, following the formulation introduced by (Bar-Shalom, Daum, & Huang, 2009). Instead of associating a single measurement with a track, JPDA computes a weighted average over multiple feasible measurements, each weighted by its likelihood of association with the current track.

Let z_i be the ith measurement in the validation gate, and β_i be its associated probability. The weighted innovation vector is computed as:

$$\bar{\mathbf{z}} = \sum_i \beta_i \, (\mathbf{z}_i - \mathbf{H}\hat{\mathbf{x}}^-)$$

The innovation covariance is altered to consider the variance of the cluster of the sensor measurements:

$$S = \frac{\sum_i \beta_i (\mathbf{H}P^-\mathbf{H}^T + R_i)}{\sum_i \beta_i}$$

The EKF gain K is followed as the typical variant:

$$K = P \cdot \mathbf{H}^T \cdot S^{-1}$$

After computing K, the state and covariance are updated as:

$$\hat{\mathbf{x}}^+ = \hat{\mathbf{x}}^- + K\bar{\mathbf{z}}$$

$$P^+ = P^- - \sum_i \beta_i K S K^T + K\left(\sum_i \beta_i (v_i v_i^T) - \overline{\mathbf{z}}\overline{\mathbf{z}}^T\right)K^T$$

Where $v_i = \mathbf{z}_i - \mathbf{H}\hat{\mathbf{x}}^-$ is the innovation for the ith observation. The second correction term accounts for the dispersion of individual innovations around the weighted mean, thereby reducing estimation bias when multiple ambiguous measurements are present.

5.2 Data association

A Joint Probabilistic Data Association (JPDA) method was adopted and integrated into the EKF framework to address the association between measurements and existing tracks. Unlike simpler methods, such as nearest-neighbour approaches, JPDA does not rely on a single best measurement. Instead, it evaluates all plausible measurement-to-track associations simultaneously and updates each track using a probabilistic combination of observations. A schematic overview of the JPDA method is illustrated in Figure 5.

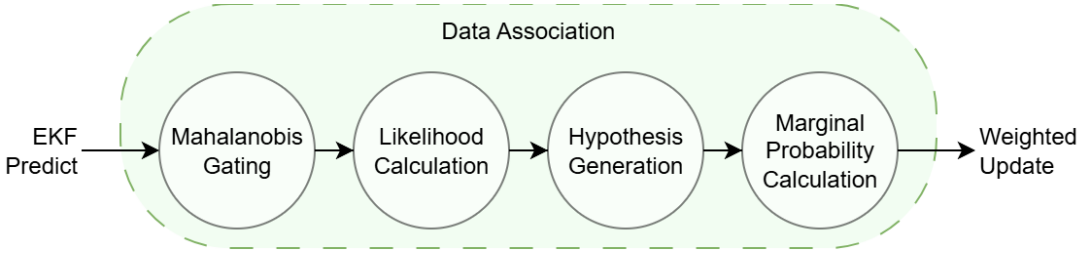


Figure 5 - Flowchart of the Joint Probabilistic Data Association approach.

Prior to evaluating associations, the set of incoming observations is reduced through a statistical compatibility test. This test is based on the discrepancy between each measurement and the predicted value obtained by propagating the EKF state through the observation model. The discrepancy is quantified using the Mahalanobis distance:

$$d_i^2 = (\mathbf{z}_i - \mathbf{H}\hat{\mathbf{x}}^-)^T S^{-1} (z_i - \mathbf{H}\hat{\mathbf{x}}^-)$$

Only measurements that satisfy the predefined threshold $d_i^2 < \gamma$, where γ is selected from the appropriate chi-squared distribution, are considered further.

For each remaining measurement-track pair, an association likelihood is evaluated. This likelihood reflects how well the observation matches the predicted measurement distribution of the EKF, which is approximated as Gaussian after linearization. The result takes the form:

$$\mathcal{L}_{i,j} = \frac{1}{\sqrt{(2\pi)^n |S|}} \exp\left(-\frac{1}{2} d^2\right)$$

And represents the probability that measurement j was generated by track i, assuming zero-mean Gaussian noise. All valid measurement-to-track assignments are enumerated, where each association hypothesis θ assigns either a single observation to each track, or no observation (missed detection). Each hypothesis is assigned a joint probability $P(\theta)$, computed as the product of:

- individual association likelihoods $\mathcal{L}_{i,j}$,
- missed detection penalties $(1 - P_D)$, and

- false alarm penalties based on a Poisson clutter model:

$$P(\theta) = \prod_{i=1}^{T} \begin{cases} P_D \cdot \mathcal{L}_{i,j}, & \text{if } \theta(i) = j \\ 1 - P_D, & \text{if } \theta(i) = 0 \end{cases} \cdot L_{FA}^{N_{FA}}$$

Where T is the number of tracks, P_D is the probability of detection, $L_{FA} = \lambda V$ is the false alarm model over the validation volume V, and N_{FA} is the number of unassigned observations. All hypotheses are normalized with:

$$\tilde{P}(\theta) = \frac{P(\theta)}{\sum_{\theta} P(\theta)}$$

The normalized hypothesis probabilities are used to compute the marginal association probabilities $\beta_{i,j}$, representing the probability that measurement j originated from the track i:

$$\beta_{i,j} = \sum_{\theta:\theta(i)=j} \tilde{P}(\theta)$$

These $\beta_{i,j}$ values are then passed to the EKF, where they are used to compute the weighted innovation vector and update the state and covariance.

5.3 Track Management

Beyond the state estimation and data association, a dedicated track management module is employed to regulate the lifecycle of tracked entities, as seen in Figure 6. In this context, each track is assumed to correspond to a single target and maintains its own state vector and covariance. The role of this module is to decide when such target representations should be initiated, maintained, or removed.

When incoming sensor measurements cannot be associated, with sufficient probability, to any existing track, they are interpreted as indications of newly appearing targets. This enables the system to introduce new target hypotheses as they enter the sensor's field of view.

Track deletion is based on both temporal and statistical criteria. If a track is not updated by a sensor measurement within a predefined time threshold, it is assumed that the target is no longer observable or has exited the monitored area and is therefore removed. Additionally, a track's covariance that grows larger than a predefined threshold indicates excessive uncertainty in the estimated target state, often caused by repeated missed detections or weak observability. These tracks are considered unreliable and discarded.

Tracks that receive valid sensor associations during the JPDA stage are updated using EKF and remain active target representations.

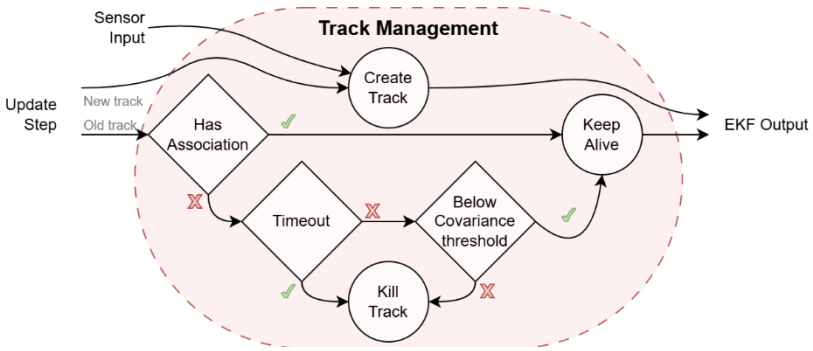


Figure 6 - Track management module flowchart.

6 Foundational Sensor Processing

Equipping the platform with a diverse set of sensors is only the first step toward achieving reliable perception and autonomy. To ensure that the information provided by these sensors can be effectively utilized in localization and multimodal fusion, their outputs must be integrated into a coherent and consistent framework. This requires aligning data streams in time, establishing a common reference across modalities, and integrating them within a unified software environment.

Equally important is the precise knowledge of each sensor's placement and configuration on the vehicle, as even small spatial or temporal discrepancies can significantly degrade perception performance. In addition, the raw output from different sensors often requires filtering, conditioning, or selection to ensure that only relevant and usable information is propagated through the perception pipeline.

This section provides an overview of the foundational processing steps that enable higher-level perception tasks. The following subsections describe the main components of this preparation layer, including registration and synchronization, sensor calibration, and sensor pre-processing.

6.1 Registration and Synchronization

The synchronization and registration architecture for the AIRSHIP perception system was previously introduced in Deliverable D6.1. For completeness, this section provides a brief overview of the underlying approach before presenting the updates and current implementation status. The perception framework is built on ROS 2, which offers native support for timestamping, data flow management, and tracking of geometric transformations through its middleware and TF2 tools. These capabilities provide a robust foundation for ensuring temporal consistency and maintaining a coherent spatial representation across all sensor modalities.

The synchronization strategy adopted in this work follows the same architectural principles defined in D6.1 and depicted in Figure 7, which outlines the different layers of the sensing suite.

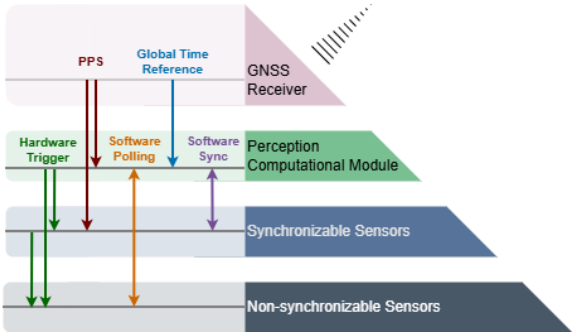


Figure 7 - Synchronization architecture.

Each sensor is integrated into this architecture according to its specific characteristics and interface requirements. Table 3 provides an overview of those that provide synchronization capability and their corresponding configuration.

Table 3 - Sensor synchronization method.

AIRSHIP sensor	Synchronization Method
LsLidar LS180S2	PTP
ARS 548 RDI	PTP
GuideNav GUIDE730	Free-running (with *data ready* pin)
ArduSimple simpleRTK3B	PPS output
Dalsa Genie Nano GigE	PTP + Software trigger
Calibir GXM	PTP + Software trigger

6.2 Sensor Calibration

The calibration of the sensing system is essential to ensure that all sensor outputs are accurate, consistent, and suitable for downstream perception tasks. In general, calibration can be divided into two complementary components: intrinsic calibration and extrinsic calibration.

Intrinsic calibration concerns the internal parameters of each sensor. Although many sensors are factory-calibrated, imaging devices require additional calibration because lenses, optical filters, or protective glass domes can introduce distortions that alter the final image.

To compensate for these effects, standard camera calibration procedures are applied. In this work, a chessboard-based method is used to estimate lens distortion and refine the projection model.

In Figure 8, the difference between the original and the undistorted camera image can be clearly observed, proving that this process is necessary, especially for image edges where the distortion is more pronounced.



Figure 8 - Original and undistorted camera image.

Extrinsic calibration defines the relative spatial relationships between sensors. When the platform geometry is precisely modelled in 3D, and sensors are mounted with tight mechanical tolerances, these transformations can be derived directly from the design data. Otherwise, dedicated calibration steps must be performed to estimate the relative poses between two or more sensors. In such procedures, all sensors involved must observe the same calibration target or scene so that their outputs can be correlated reliably. A common example is the calibration

of a stereo camera pair, see Figure 9, where even minor misalignments can lead to significant geometric inconsistencies, requiring correction to ensure accurate depth estimation.

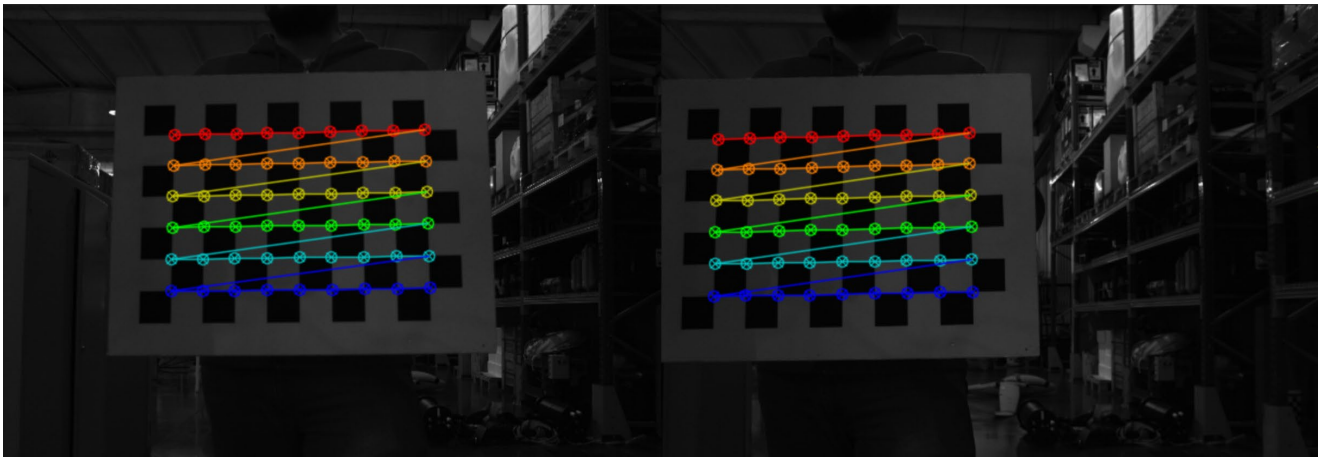


Figure 9 - Stereo pair calibration.

If the airframe exhibits structural deformation, such as torsion or bending, during more turbulent flight stages, additional methods based on SIFT, SURF, or ORB can be employed for baseline correction.

6.3 Pre-processing

The r output produced by the sensing suite is not always directly suitable for real-time perception. In many cases, the volume, noise level, or structure of the data can make it impractical to process at full resolution. For example, range sensors such as LiDAR's can generate dense point clouds that would impose a substantial computational burden if used without any reduction or filtering, particularly in the context of real-time target tracking or fusion algorithms.

To ensure that the perception pipeline remains computationally efficient while preserving the information necessary for reliable operation, pre-processing steps must be applied to selected sensor outputs. These steps simplify data complexity, eliminate irrelevant or low-quality measurements, and restructure the data into more manageable representations.

This subsection describes the pre-processing methods employed in this work, starting with range-sensor data and extending to other modalities.

6.3.1 Range Sensor Clustering

To reduce the density of LiDAR point cloud data and extract meaningful structures, clustering methods are employed. Among the commonly used algorithms, k-means, DBSCAN, and hierarchical clustering represent three distinct approaches with different advantages depending on the characteristics of the data.

- Hierarchical clustering builds a tree-like structure of nested clusters by either progressively merging nearby clusters or recursively splitting larger ones. Although it offers a flexible multiscale representation, this method was excluded due to its significantly higher computational demands, which are unsuitable for real-time processing of dense point clouds.

- DBSCAN operates on a density-based principle, grouping points that lie within dense neighborhoods while labeling sparse points as noise. It can identify clusters of arbitrary shape and does not require a predefined number of clusters, making it robust for LiDAR data, see Figure 10.

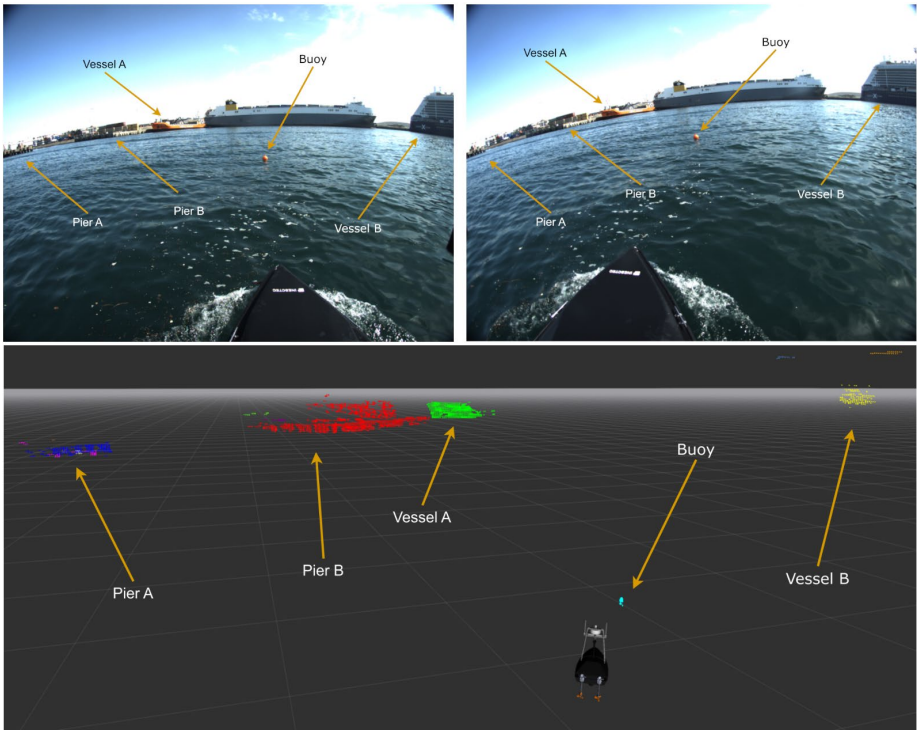


Figure 10 - LiDAR clustering output.

- Euclidean clustering groups data points based on their spatial proximity by forming clusters when the Euclidean distance between neighbouring points falls below a specified threshold. This makes it well-suited for sensors such as RADAR, see Figure 11, where detections are sparse, and clusters are naturally defined by physical distance rather than centroid-based partitioning.

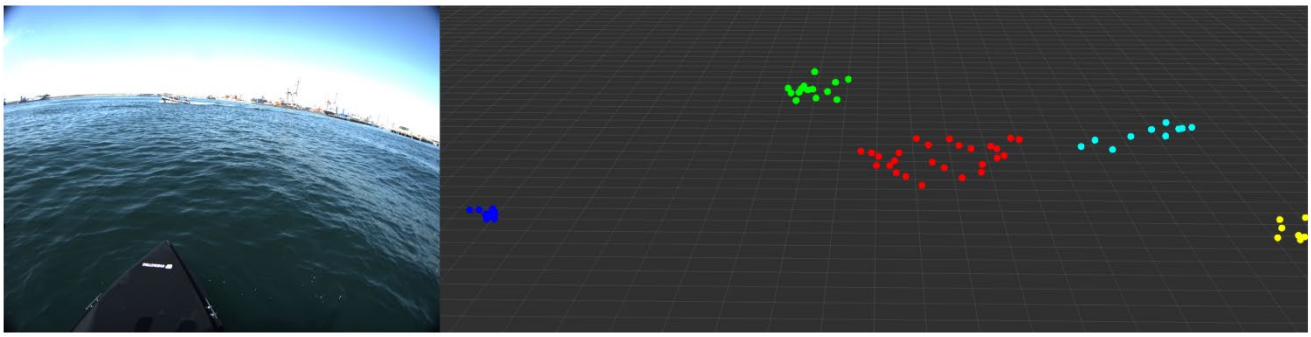


Figure 11 - RADAR clustering output.

Consequently, the range sensor data can be distilled into a set of candidate targets characterized by their centroid locations and spatial extents. This abstraction replaces dense point clouds with a lightweight representation, lowering computational overhead while inherently filtering out noise.

6.3.2 Visible Image Processing

EO cameras provide rich visual information but do not directly supply the 3D measurements required by the target-tracking filter. To obtain spatially meaningful observations, the perception system employs a stereo pair of EO cameras, enabling depth estimation through stereo triangulation. This configuration enables the extraction of a 3D point for every pixel in the image, thereby producing a dense depth map that is aligned with the visual scene.

Processing the full set of reconstructed 3D points, however, would be computationally inefficient and unnecessary for real-time operation. To constrain the data flow and focus only on relevant regions, an object-detection model is first applied to the image stream. In this work, inference is performed using NVIDIA-based PC, which detects classes such as vessels and buoys. The output of this inference step identifies image regions that are likely to correspond to potential targets. See an example in Figure 12.



Figure 12 - Target detection on the input image.

Once a region of interest is detected, the corresponding depth values from the stereo reconstruction are extracted. This association enables the computation of the 3D position of the detected object by mapping its image-space bounding box to the underlying 3D point cloud. The resulting spatial measurement, represented by the centroid of the selected 3D points, can then be passed to the tracking filter as an observation.

An example of this process is illustrated in Figure 13, where a vessel detected in the EO image is matched with its corresponding 3D structure in the stereo-derived point cloud.

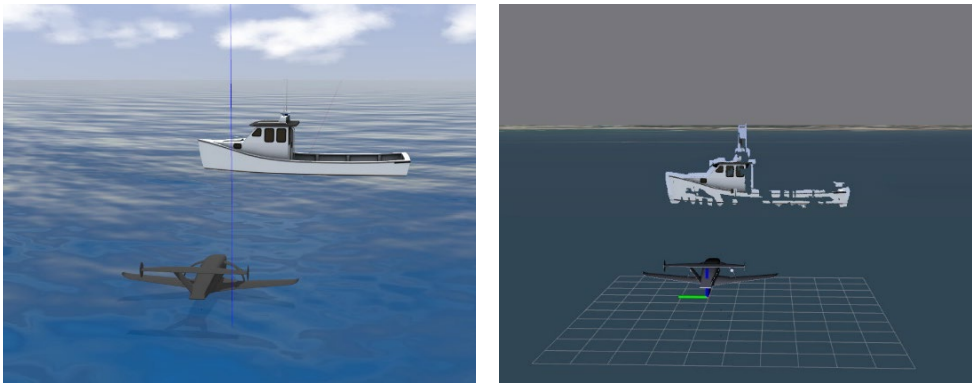


Figure 13 - Point cloud reconstruction from a stereo pair.

6.3.3 Thermal Image Processing

Detecting obstacles on the sea surface (for example, a small buoy) with a thermal camera mounted on an unmanned WIG vehicle is a classic "low-contrast, high-clutter" maritime vision problem. Even though thermal sensing can reduce dependence on visible illumination and mitigate glare compared to EO cameras, the ocean presents a challenging background: it is highly textured (with whitecaps and wakes), dynamically changing, and its apparent temperature can vary with wind, humidity, and viewing geometry. On top of that, a WIG platform flies very low, so it is expected strong ego-motion, frequent horizon occlusions, spray, and rapid perspective changes. These characteristics do not allow the use of traditional frame-by-frame thresholding.

A first family of solutions relies on classical detection pipelines. Common steps include video stabilization (IMU-assisted or purely visual), horizon line detection, sea-background suppression, and the application of algorithms for possible obstacle detection and tracking. Obstacle extraction can start with adaptive local contrast measures (e.g., top-hat filtering, local z-score, CFAR-like logic) rather than a global threshold, because the sea's thermal field is nonstationary. After extracting "hot" or "cold" blobs (depending on possible obstacle temperature relative to water), using simple geometric constraints (size range in pixels, compactness, temporal persistence), these detections can be used as input to a multi-target tracker to reject transient wave crests.

A second family of solutions uses learning-based detectors (Bovcon & Kristan, 2021), (Chen, Liu, & Achuthan, 2021), typically a Convolutional Neural Network (CNN) trained on thermal frames to output obstacle candidates. These methods can learn subtle contextual cues and can outperform handcrafted features in cluttered conditions. Performance can be further improved through data augmentation, which incorporates various sea states, altitudes, and sensor settings. For the long-range small-target case, a small-object-aware approach is required: higher input resolution, multi-scale feature pyramids, super-resolution or feature upsampling, and training losses that don't ignore tiny boxes. The main advantage is robustness across environmental variability once properly trained. The disadvantages include the need for representative labelled data (thermal maritime datasets are rarely "ready-made"), the risk of performance collapse outside the training distribution (such as new sea states or different obstacle materials), and higher computational costs. For a WIG vehicle with tight SWaP constraints, a hybrid approach can be an effective solution: a lightweight candidate generator to reduce the search area, combined with a compact classifier for verification.

The CMO (close-minus-open) method (Almeida, 2009) is particularly relevant as a pragmatic background-suppression technique in thermal maritime scenes. Conceptually, two images are computed: a "close" version where local neighbourhoods are emphasized (preserving small, sharp anomalies), and an "open" version that represents the slowly varying background (suppressing small bright or dark features). Subtracting the open image from the close image tends to cancel broad sea temperature gradients while preserving compact targets, which is exactly the desired behaviour given that the target is expected to be at long range and potentially small, occupying only a few pixels in the image. In practice, "open" is often implemented with a morphological opening (erosion followed by dilation) or a large-scale smoothing operator; "close" can be a closing, a smaller-scale smoothing, or a local-contrast-enhancing transform depending on whether the obstacle appears hotter or colder than water. The strength of CMO is that it is simple, fast, and explainable, and it naturally acts like a band-pass filter tuned to the obstacle's expected spatial scale. Its weakness is that wavelets, foam patches, and sun-warmed ripples can also produce localized anomalies at similar scales - so CMO usually works best when combined with temporal consistency checks and stabilization (so subtraction doesn't amplify motion blur and jitter).

After identifying an object in the thermal image, with the purpose of enabling its tracking using the EKF, two possible approaches to estimating its 3D position were considered: The first approach is analogous to the method used with EO cameras. In this case, stereo reconstruction is performed using one EO camera and one thermal camera, rather than two EO cameras. However, this setup is limited by the EO camera, since its output is strongly influenced by illumination conditions.

The second approach relies on LiDAR or RADAR data. By projecting the range sensor point cloud onto the thermal image, each pixel in the image can be associated with a corresponding depth value. An example of this process is shown in Figure 14 where, on top of the chessboard used for thermal camera calibration, the LiDAR point cloud is projected, allowing for the determination of the chessboard's 3D position.

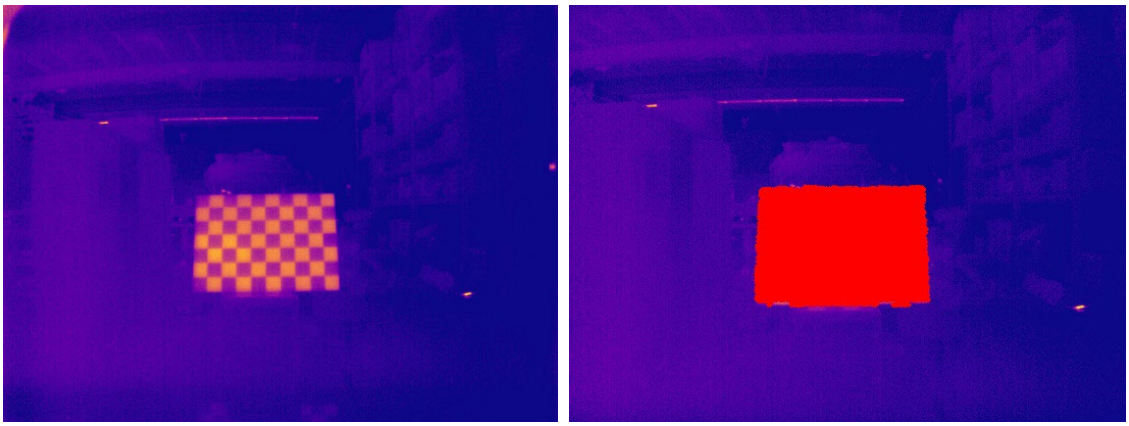


Figure 14 – 3D LiDAR point cloud projected on the 2D thermographic image.

Both methods enable a detected target in the thermal image to be directly associated with a three-dimensional position, which can be used as input to the target-tracking filter.

Figure 15 illustrates the overall thermal imaging processing framework. Starting from the thermal image input, different object detection approaches can be applied depending on the operational environment. Classical pipelines are suitable for large obstacles, while the CMO method is effective for long-range and small targets. Learning-based approaches are employed when higher-level semantic information is required. The detected objects are expressed in image coordinates $[u, v]$ and forwarded to the 3D positioning module. Depth estimation is performed either through stereo reconstruction or by projecting range sensor point clouds onto the thermal image, resulting in a three-dimensional target position $[x, y, z]$ for subsequent tracking and navigation.

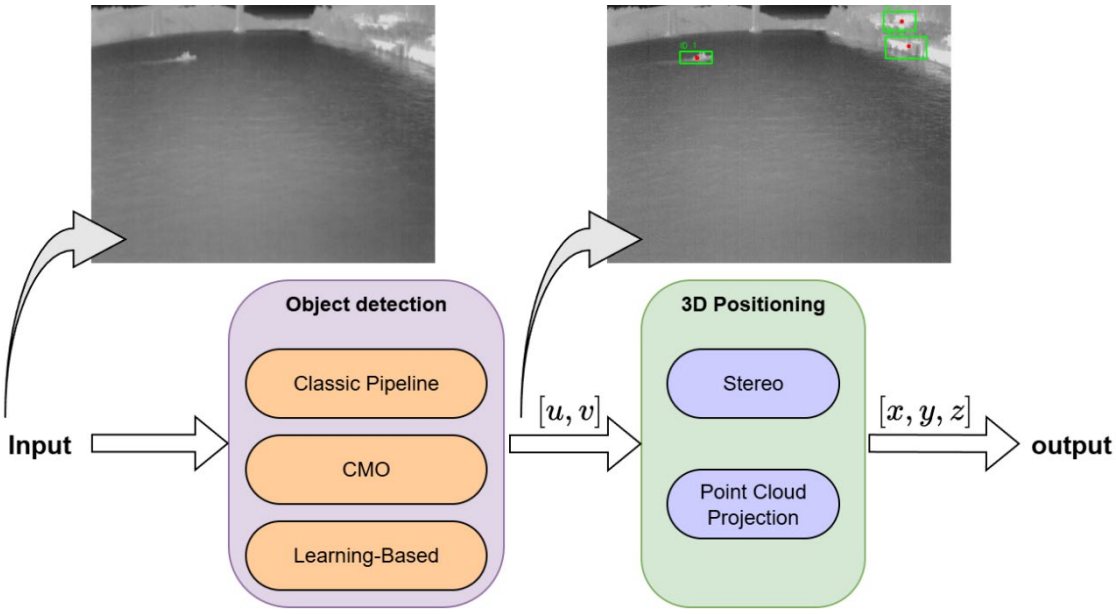


Figure 15 - Thermal image processing architecture.

7 Tests and Results

Existing simulators are becoming increasingly realistic. However, they still struggle to capture all the conditions present in real-world environments. Regarding the perception system, several factors can compromise its effectiveness, including vibrations along the structures where the sensors are mounted (affecting extrinsic calibration), sunlight reflections on the water, fog, wind (which generates low-profile waves), and ambient temperature. For example, in a simulated environment, it cannot be determined whether the processing load will cause the CPU temperature to rise to the point of triggering a protective system shutdown.

On the other hand, in real-world conditions, using either the test platform (PORTUS USV) or even a support boat, it was unable to reach the cruising speed expected by the model - a concern raised during the consortium meeting in Tampere in May 2025.

With this in mind, the tests were conducted both in simulation and in the field.

7.1 Field tests

The real-world tests were conducted in the manoeuvring basin of the Port of Leixões, Portugal. The test platform used was the previously introduced PORTUS USV Figure 16, equipped with the payload designed for AIRSHIP. As an obstacle, EPISEA (INESCTEC's support boat) was used, which was equipped with a GNSS logger (without RTK) to evaluate the results produced by the developed perception system.

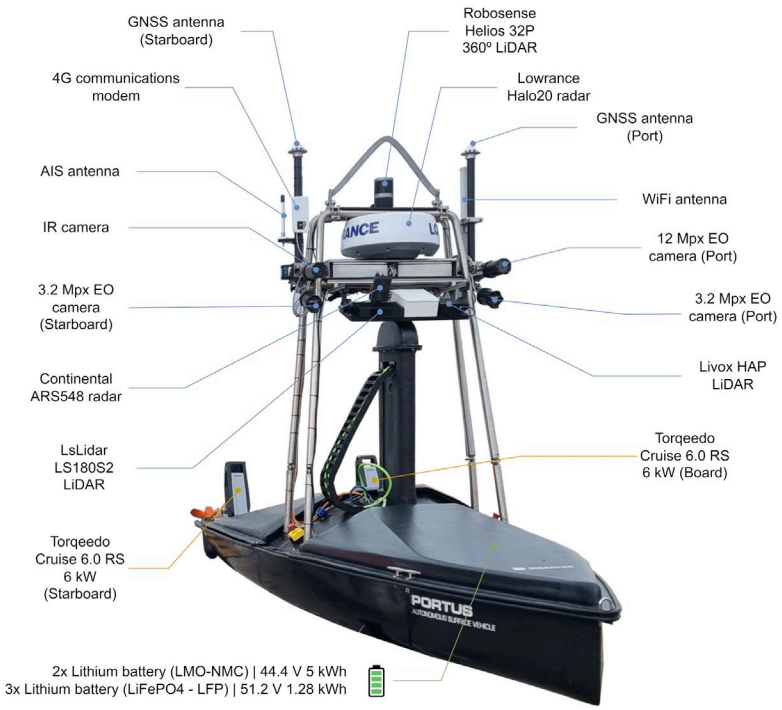


Figure 16 - PORTUS USV.

Although the area is sheltered and protected from waves and wind, the movement of other vessels generates significant wake, causing the vehicle's attitude to be noticeably disturbed - particularly in roll and pitch, depending on the angle of incidence between the vehicle and the waves. Even under these oscillatory conditions,

the tracking system was able to follow the obstacle satisfactorily. A proper definition of the translations and rotations between the different sensors (extrinsic calibration) is essential for achieving good results, even when the vehicle is subjected to disturbances. Regarding the ground-truth position data for the obstacle, considerable jumps in altitude were observed due to the absence of RTK, making a fair three-dimensional analysis infeasible. For this reason, the evaluation focused on two-dimensional analysis, assuming no variations in altitude.

7.1.1 Field Test Results

Due to technical issues in the communication interface of the perception system, the imaging sensors, i.e., cameras and thermographic cameras, did not provide reliable data. As such, the perception architecture was adapted to operate exclusively with LiDAR and RADAR information, ensuring the system could still perform robust target tracking with the available sensors.

Figure 17 presents the restructured perception system used in the field mission. The most significant modification was the inclusion of a multi-modal fusion block (hybrid fusion), designed to integrate the LiDAR and RADAR measurements before feeding them into the Kalman filter.

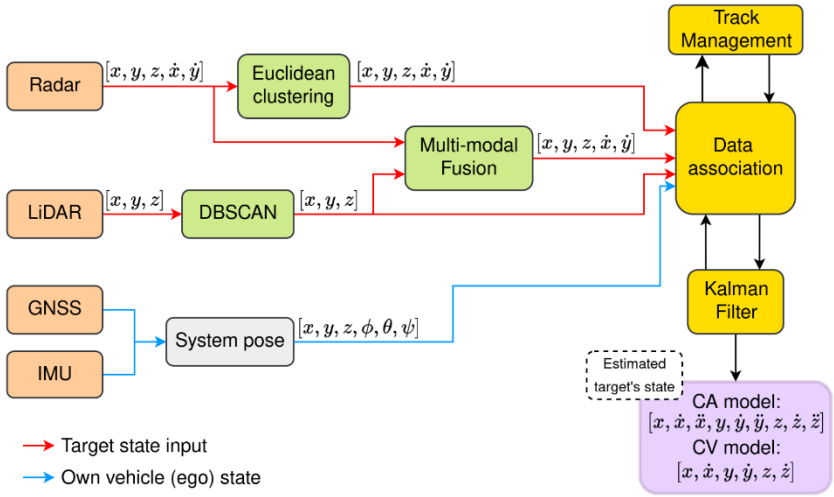


Figure 17 - Perception system's diagram used in field missions.

The main objective of the fusion module is to leverage the complementary strengths of both sensors. Specifically, LiDAR offers high spatial resolution for accurate target positioning, while the RADAR provides reliable velocity measurements. By combining both sources, the system maintains LiDAR's detailed spatial perception and augments it with the dynamic information obtained from RADAR.

In practice, this fusion process involves associating RADAR detections with the corresponding LiDAR point clusters and computing the average velocity components in the x and y axis for each matched cluster, as observed in Figure 18.

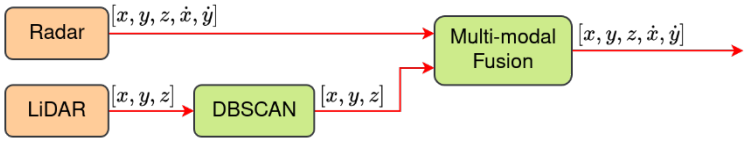


Figure 18 - Multi-modal fusion diagram.

Three configurations were tested within the Kalman filter framework: LiDAR-only, RADAR-only and the multi-modal fusion. Each configuration was further evaluated using two distinct motion models (CV and CA), resulting in a total of six experimental setups for performance comparison.

The perception system data and EPISEA's GNSS were logged using ROS2 bag files on separate computers. EPISEA served as a ground truth reference and followed a trajectory starting from a relatively distant position and gradually approach the sensor platform ASV PORTUS. After the experiments, data was synchronized offline by aligning timestamps and Kalman outputs were extracted from LiDAR only, RADAR-only, and fused configurations for performance analysis.

Figure 19 and Figure 20 compare the tracked trajectories using CV and CA Kalman filters for the previously mentioned configurations.

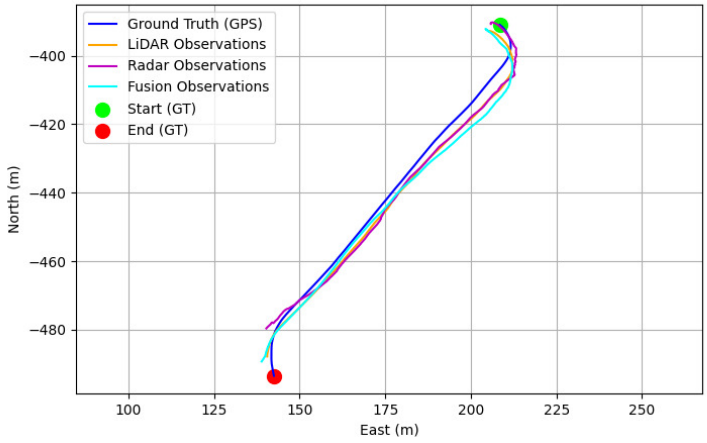


Figure 19 - Different Kalman configurations trajectories using CV model compared to ground-truth.

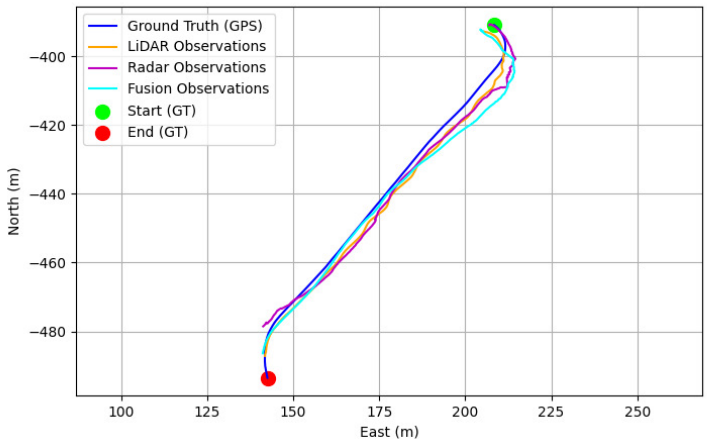


Figure 20 - Different Kalman configurations trajectories using CA model compared to ground-truth.

The results summarized in Table 4 and Table 5 show the 2D and 3D Euclidean position error statistics for each combination of motion model and sensor configuration, respectively.

Table 4 - 2D Euclidean Error Statistics

Motion Model	Filter Configuration	Mean Error (m)	Median Error (m)	Std. dev (m)
CV	LiDAR-only	0.11133	0.11579	0.03840
	RADAR-only	0.20038	0.20969	0.06352
	Fused	0.13100	0.11732	0.04737
CA	LiDAR-only	0.11133	0.11664	0.03774
	RADAR-only	0.21279	0.22806	0.05593
	Fused	0.14669	0.13334	0.05345

Table 5 - 3D Euclidean Error Statistics

Motion Model	Filter Configuration	Mean Error (m)	Median Error (m)	Std. dev (m)
CV	LiDAR-only	2.32653	2.37048	0.85788
	RADAR-only	4.74427	4.16854	1.32136
	Fused	2.42037	2.30885	0.64670
CA	LiDAR-only	2.32247	2.23437	0.84276
	RADAR-only	4.76351	4.28032	1.29744
	Fused	2.41939	2.36678	0.61905

For 2D tracking performance, the LiDAR-only configuration achieved the lowest mean error under the CA model, likely due to the high spatial resolution and accuracy. However, the fused configuration showed competitive performance with the CA and CV model, while providing enhanced robustness through velocity information from radar. RADAR-only configurations consistently exhibited the highest errors, with mean values exceeding 0.20 meters in both motion models, reflecting the inherent limitations of RADAR in terms of spatial resolution.

The 3D tracking results present a different picture, with all configurations showing significantly higher error magnitude. The best 3D performance was achieved by using LiDAR-only configuration under the CA model. Notably, the fused configuration maintained relatively reasonable accuracy, demonstrating only marginal degradation compared to the LiDAR-only performance while gaining the benefit of velocity estimation.

Figure 21 and Figure 22 illustrate the temporal evolution of tracking throughout the experimental scenario. A clear trend emerges where 3D tracking errors decrease as EPISEA approaches the sensor platform, particularly evident in the final seconds of the trajectory. This improvement can be attributed to several factors, namely: reduced angular uncertainty at closer ranges, improved LiDAR and radar point cloud density on the target, and mitigation of water surface reflections that become less pronounced at better scanning angles. Additionally, part of the initial degradation can be attributed to the track initialization stage of the filter. At the beginning of each run, tracks are instantiated with limited prior information, and the initial covariance is intentionally set conservatively to accommodate measurement uncertainty. As new detections are incorporated, the filter converges toward the true state, resulting in reduced errors after the first few seconds of tracking.

The 2D error profiles remain relatively stable throughout the tracking period. Standard deviation values in both tables indicate that the fused approach provides more stable estimates, with reduced variance compared to individual sensor configurations.

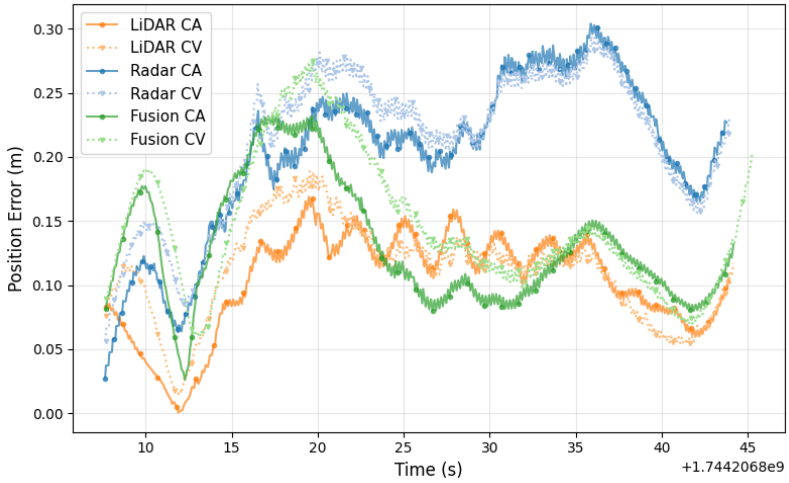


Figure 21 - 2D Euclidean error of Kalman output using LiDAR-only, RADAR-only, and fused measurements for CA and CV motion models.

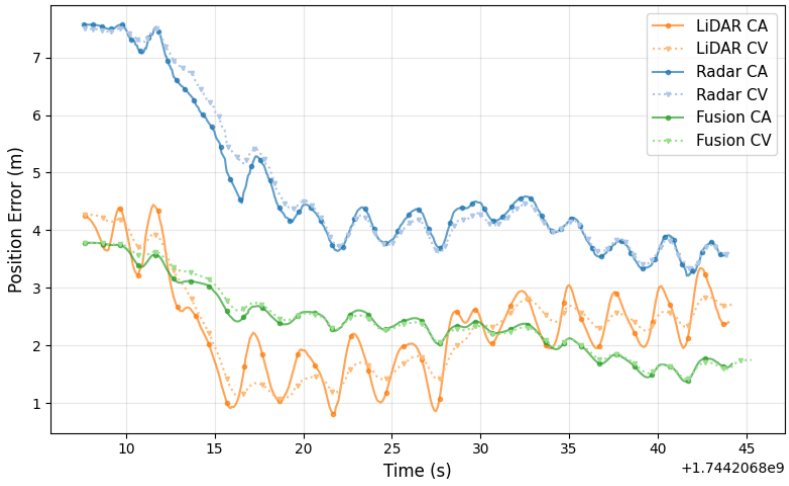


Figure 22 - 3D Euclidean error of Kalman output using LiDAR-only, RADAR-only, and fused measurements for CA and CV motion models.

Overall, both motion models delivered nearly identical accuracy across all sensor configurations. The fused data setup consistently matched the LiDAR-only configuration performance in both 2D and 3D domains while showcasing lower variance, confirming improved stability. The results indicate that, within the testbed conditions, the choice between the two motion models has a limited effect on the tracking accuracy. The inclusion of radar information mainly contributes to smoother and more robust estimates, particularly in conditions where LiDAR returns are partially degraded.

Although the filter couldn't be fully validated due to technical issues with the imaging sensors, it was possible to validate the robustness of the perception system, demonstrating that it is capable of reliable tracking even in situations where cameras are damaged. This emphasizes the inherent robustness and redundancy benefits of a multi-modal perception architecture.

7.2 Simulation

Although real-world data is generally more valuable as it captures natural sensor noise, environmental variability, and unmodeled physical effects, it also has practical limitations. In the conducted field tests, the testbed platform was PORTUS USV, which does not reach the high speeds characteristic of a WIG vehicle. While the USV-based experiments still provide representative data thanks to the shared marine environment and near-surface operating conditions, they cannot fully reproduce the dynamic scenarios associated with high-speed flight just above the water. Furthermore, real-world testing is inherently constrained by the number and type of targets physically present in the environment, limiting the variety and density of scenarios that can be evaluated.

To address these gaps, simulation was employed to assess the robustness and performance of the developed perception system under higher-speed conditions and more diverse target configurations. Using the Gazebo Classic simulator, it was possible to create controlled, repeatable environments with multiple targets of varying trajectories and behaviours, thereby enabling testing of the system. A detailed description of the simulation setup and configuration is provided in the following sections.

7.2.1 Simulation Scenario

Given the intended operational context of a WIG craft, characterized by near-surface, high-speed motion over open water, an ocean environment was selected for the simulation scenario.

The WIG vehicle model used in the simulations is shown in Figure 23. The model was equipped with primary sensors relevant to the validation objectives: a LiDAR, RADAR, and a stereo camera pair. Other onboard sensors available on the real platform were intentionally excluded, either because their simulated data would not be sufficiently meaningful compared to real-world measurements or because the same information could be obtained directly from the simulation environment, where full control over the world state is available. The spatial placement of the included sensors is also illustrated in Figure 23, where the LiDAR is highlighted in blue under the vehicle nose, the radar in yellow, and the stereo cameras in green and red to distinguish between them.

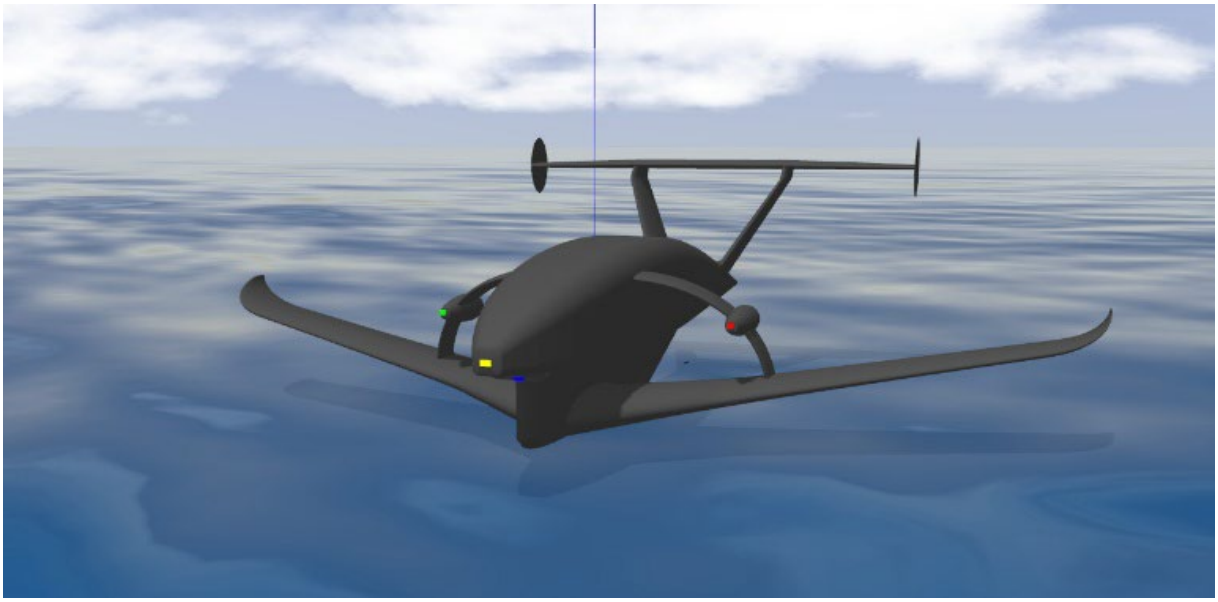


Figure 23 - WIG model used in the simulation and its sensors.

To evaluate the system's multi-target detection and tracking capabilities, two types of maritime targets were introduced into the environment: a vessel (14 m x 5 m x 5 m) and a buoy (Ø 0.5 m x 1.2 m), see Figure 24. These objects were duplicated and arranged in various positions to emulate some target density and more diverse encounter geometries. Such configurations allowed the perception system to be tested under conditions more challenging than those available during field experiments, particularly in scenarios requiring rapid detection and consistent tracking at higher speeds. Furthermore, the vessels were assigned different velocities, ranging from slow to fast, to validate the system's performance limits and ensure robustness across a wide range of target dynamics.

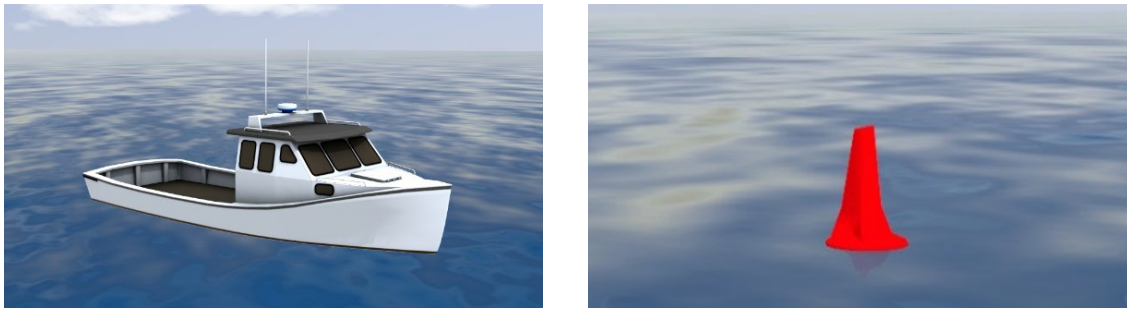


Figure 24 - Obstacles used in simulations.

Two distinct trajectories were defined for the WIG model, as shown in Figure 25. The first trajectory, referred here as the straight path trajectory, follows a lateral fly-by pattern in which the vehicle passes the targets at a distance, enabling evaluation of long-range detection and classification performance. The second trajectory, referred to as the loitering path, brings the vehicle into close proximity with the targets, simulating near-collision scenarios and stressing the system's ability to maintain stable tracking and timely obstacle awareness during rapid approach. Together, these paths provide a comprehensive assessment of perception robustness across varying encounter distances and dynamics.



Figure 25 - WIG model trajectory in the simulation. On the left, the loitering trajectory, and on the right, the straight trajectory.

To complete the description of the simulation environment, the characteristics of the virtual sensors were defined to approximate those of their real-world counterparts as closely as possible, while preserving computational feasibility. Minor differences were necessary due to limitations of the simulation engine and the need to balance performance with accuracy. Among the sensors, the LiDAR was the most affected by these constraints. The real LiDAR is a high-performance device capable of generating millions of points per second,

making it computationally demanding to emulate accurately. Attempting to reproduce such data density in Gazebo led to severe computational slowdowns and instability, ultimately forcing a reduction in point cloud density and angular resolution. Despite the simplification, the LiDAR still provided rich spatial information, enabling effective testing for tracking and estimation performance.

Table 6 summarizes the main disparities between the simulated and physical sensors to be used on the real WIG platform. These deviations primarily concern FOV, resolution, and data rate parameters, which can influence detection range, spatial precision, and overall perception fidelity. Nonetheless, the selected configurations were found to be sufficient for replicating realistic behaviour for the Kalman filter validation.

Table 6 - Comparison between simulated and real sensors.

Sensor Modality	Simulated Sensor specs.	Real Sensor specs.
Camera	30 fps, HFOV 80°, VFOV 69°	38 fps, HFOV 103°, VFOV 67°
LiDAR	10 fps, HFOV 120°, VFOV 25° 0.55° horizontal resolution 0.5° vertical resolution	10fps, HFOV 120°, VFOV 25° 0.13° horizontal resolution 0.129° vertical resolution
Radar	20 Hz, HFOV 100°, VFOV 25° 0.9° horizontal resolution 0.25° vertical resolution	20 Hz, HFOV 120°, VFOV 28° 0.5° horizontal resolution 0.1° elevation resolution

7.2.2 Sensor-level target detections

Before analysing the Kalman filter estimation results, it is necessary to examine how each simulated sensor perceives and detects the surrounding obstacles. This section presents the individual detection outputs for the stereo camera pair, LiDAR, and RADAR sensors.

7.2.2.1 Stereo Camera detections

The stereo camera pair constitutes the system's optical sensing component and provides depth information through binocular vision. With a fixed baseline of 1.52 meters, the cameras capture synchronized left and right images that are processed to generate a dense 3D point cloud of the environment. Depth estimation was performed using the OpenCV Semi-Global Matching (SGM) algorithm, selected for its balance between accuracy and computational efficiency. An example of the resulting point clouds for both static and dynamic conditions is shown in Figure 26 and Figure 27, respectively, where the impact of motion on point cloud consistency can be observed.

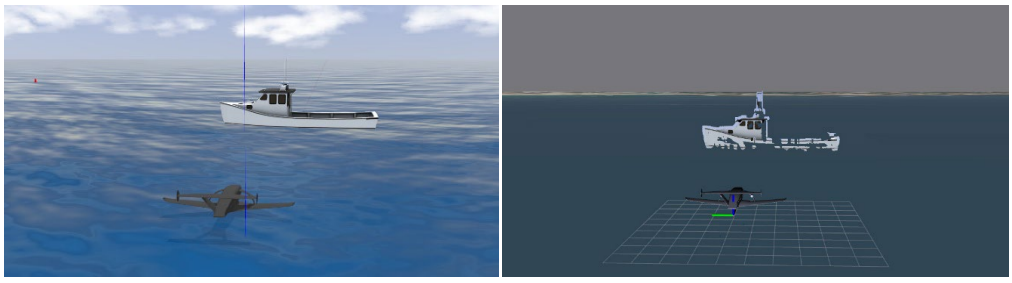


Figure 26 - Point cloud generated by stereo camera pair when the WIG is static.



Figure 27 - Point cloud generated by stereo camera pair when WIG is in motion.

The generated point clouds reveal the scene geometry around the vehicle, capturing sufficient spatial detail to distinguish primary obstacles such as vessels and buoys. When the WIG is in motion, the point clouds exhibit increased variability due to changing viewing angles. Depending on the camera orientation, certain regions of the surface are reconstructed more clearly, while others appear sparse or missing due to poor correspondence between the stereo images. Despite these variations, the overall structure of the scene remains coherent enough for effective obstacle segmentation.

To isolate the targets further, a Euclidean clustering algorithm was applied. Prior to the clustering, the point cloud was filtered along the z-axis to remove points associated with the water surface. The result of this process is shown in Figure 28, where distinct clusters corresponding to separate obstacles are highlighted using different colours.



Figure 28 - Example of clustered point cloud showing obstacle detections. Different colours represent distinct targets.

7.2.2.2 LiDAR detections

The LiDAR sensor provides a direct 3D measurement of the surrounding environment, eliminating the need for stereo triangulation. With an effective range of 250 meters, it delivers a spatially rich point cloud suitable for long-range obstacle detection and tracking. An example of the sensor output is shown in Figure 29, illustrating how the LiDAR accurately captures the geometry and position of nearby targets, even at extended distances.

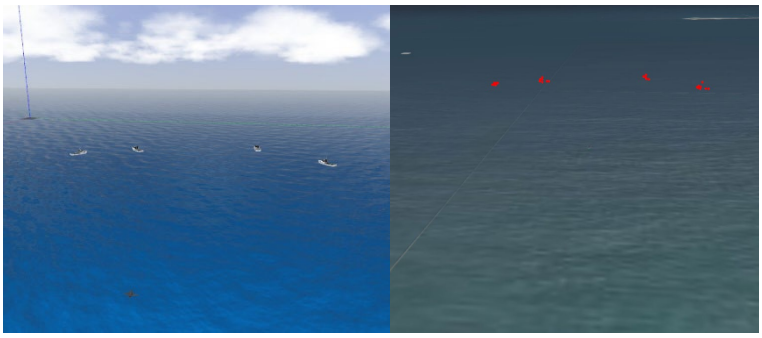


Figure 29 - Camera image (left) and raw LiDAR detections (right).

Due to its range and scanning characteristics, the LiDAR provides a consistent spatial representation across different viewing angles and motion states. However, given the reduced angular resolution used in the simulation (relative to the real sensor), fine surface details are less defined, particularly for small targets such as buoys. Despite this, the LiDAR provides accurate obstacle localization for the perception pipeline.

To identify and separate individual obstacles within the LiDAR point cloud, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was employed. The output of this clustering step is shown in Figure 30, where each target is visualized as a distinct colour cluster alongside the corresponding camera image.

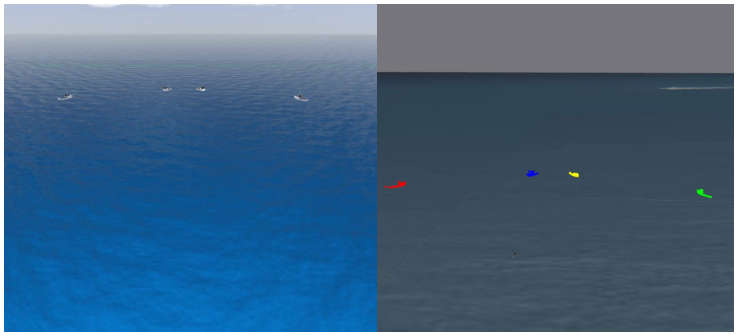


Figure 30 - Camera image (left) and LiDAR detections (right). Different colours represent distinct targets.

7.2.2.3 RADAR detections

The radar complements the optical and LiDAR sensors by providing robust detections at long range, around 1500 meters, even under conditions where visibility may be degraded. An example of these raw detections is shown in Figure 31, where the RADAR points are presented alongside the camera image, illustrating how vessels appear as sparse but clearly distinguishable returns.

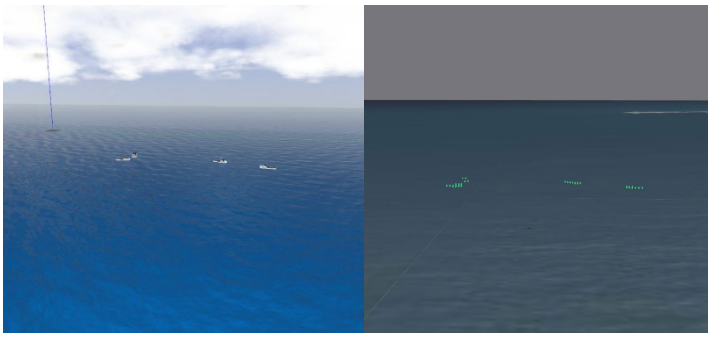


Figure 31 - Camera image (left) and raw RADAR detections (right).

Unlike the dense point clouds generated by the stereo cameras or LiDAR, radar detections tend to be fewer and more irregularly distributed, reflecting the sensor's lower angular resolution and dependence on target RADAR cross-section. Nevertheless, these measurements are highly valuable for confirming the presence of obstacles early on due to the large range, particularly for larger targets.

To group RADAR detections into individual targets, a Euclidean clustering algorithm was applied to each RADAR scan. The method associates nearby detections into coherent clusters, each corresponding to a potential target. The result of this process is shown in Figure 32, where the clustered radar detections are visualized using different colours, alongside the corresponding camera image.

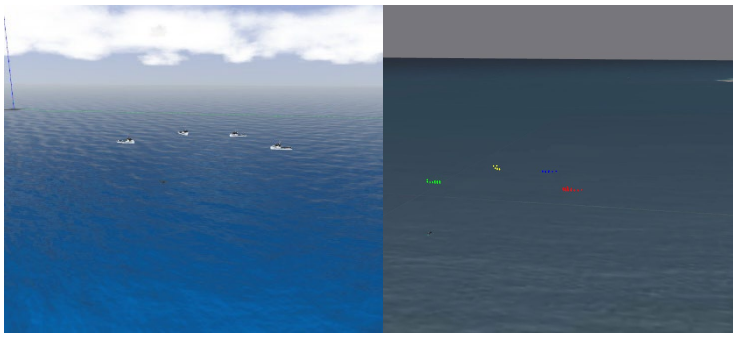


Figure 32 - Camera image (left) and RADAR detections after Euclidean clustering (right). Different colours represent distinct targets.

7.2.3 Simulation Results

Two important considerations must be taken into account when analysing the results: the ground truth refers to the centre of the obstacle (the vessel or the buoy), while the sensors observe only the external surface of the obstacle(s). Consequently, since the dimensions of the obstacle(s) are unknown, there will always be a depth offset (Δd), see Figure 33, between the observed position and the "real" position, which in turn introduces an error in the estimated position. Furthermore, this offset depends on the viewing angle, meaning that only with prior knowledge of the obstacle's shape and dimensions would it be possible to compensate for this discrepancy, see Figure 34.

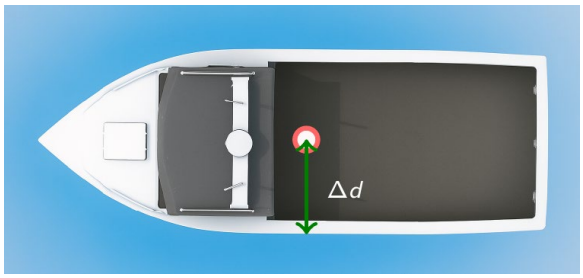


Figure 33 - Depth offset between the obstacle centre and the possible detection centroid.

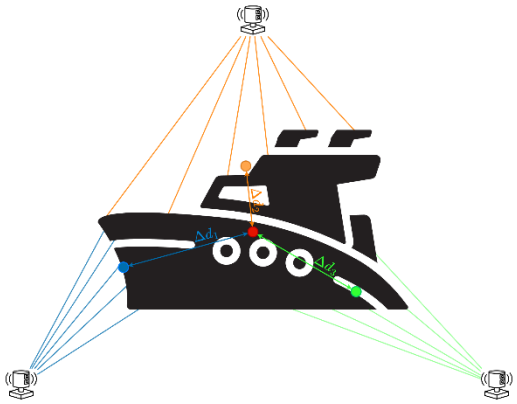


Figure 34 - Visual representation of the centroid discrepancy depending on the viewing angle of the target

To evaluate the performance of the tracking system, two trajectory scenarios were simulated: a straight path and an orbital trajectory of the WIG vehicle. These flight patterns were selected to assess the Kalman filter's ability to maintain stable target estimation under distinct viewing geometries and dynamic conditions.

In both scenarios, four vessel targets were used to test multi-target tracking performance. The targets followed circular trajectories, with two vessels moving along slower, larger circular paths (one clockwise and the other counterclockwise), and two executing faster, tighter rotations in opposite directions. All vehicle trajectories can be observed in the Figure 35. This setup offered a diverse set of relative motions and interaction distances, without collisions with each other, ensuring the filter was tested under both gradual and rapidly changing line-of-sight conditions.

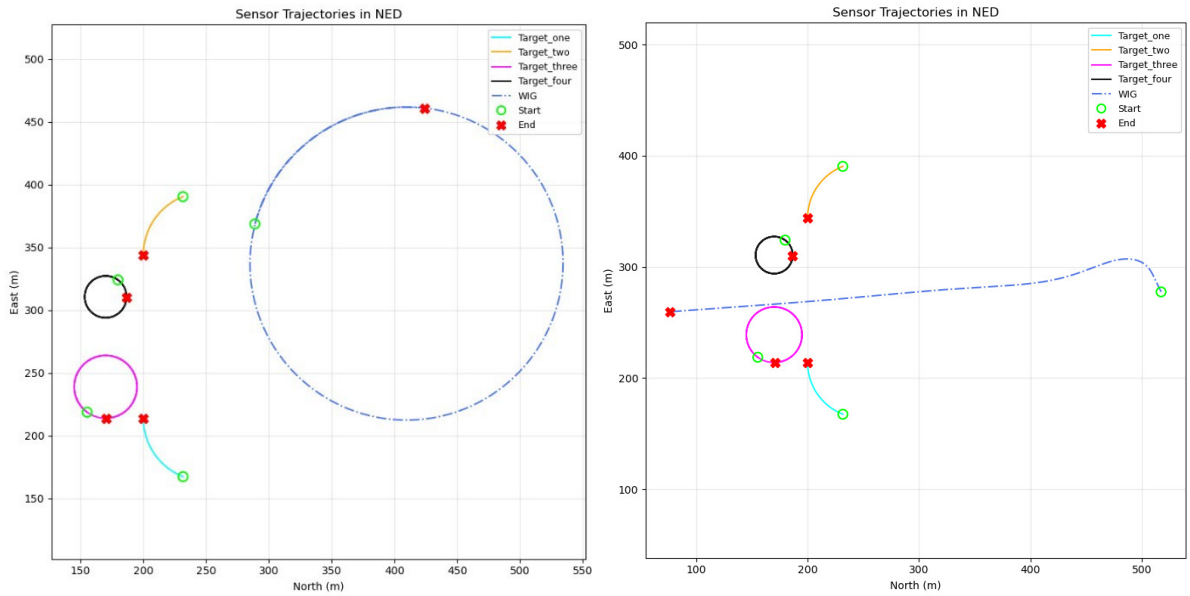


Figure 35 - Ground truth trajectories of the four vessel targets and WiG in NED coordinates

Given the physical size of the vessel targets, a positioning error on the order of meters is expected between the ground truth (target centroid) and the sensor-measured detection (cluster centroid). Depending on the WIG's position and orientation, the visible portion of each vessel changes, causing small shifts in the detected centroid position across time. To account for this, the filter model incorporated a slightly larger covariance in the measurement update, encompassing the variability between observed and true target centroids.

For the straight path trajectory, the WIG performed a fly-by manoeuvre at a relatively constant attitude and altitude (2.5 meters) with speed around 25 m/s, passing alongside the group of vessels. Figure 36 presents the comparison between the ground truth and the Kalman filter estimated trajectories for each of the four targets, shown in the NED coordinate frame.

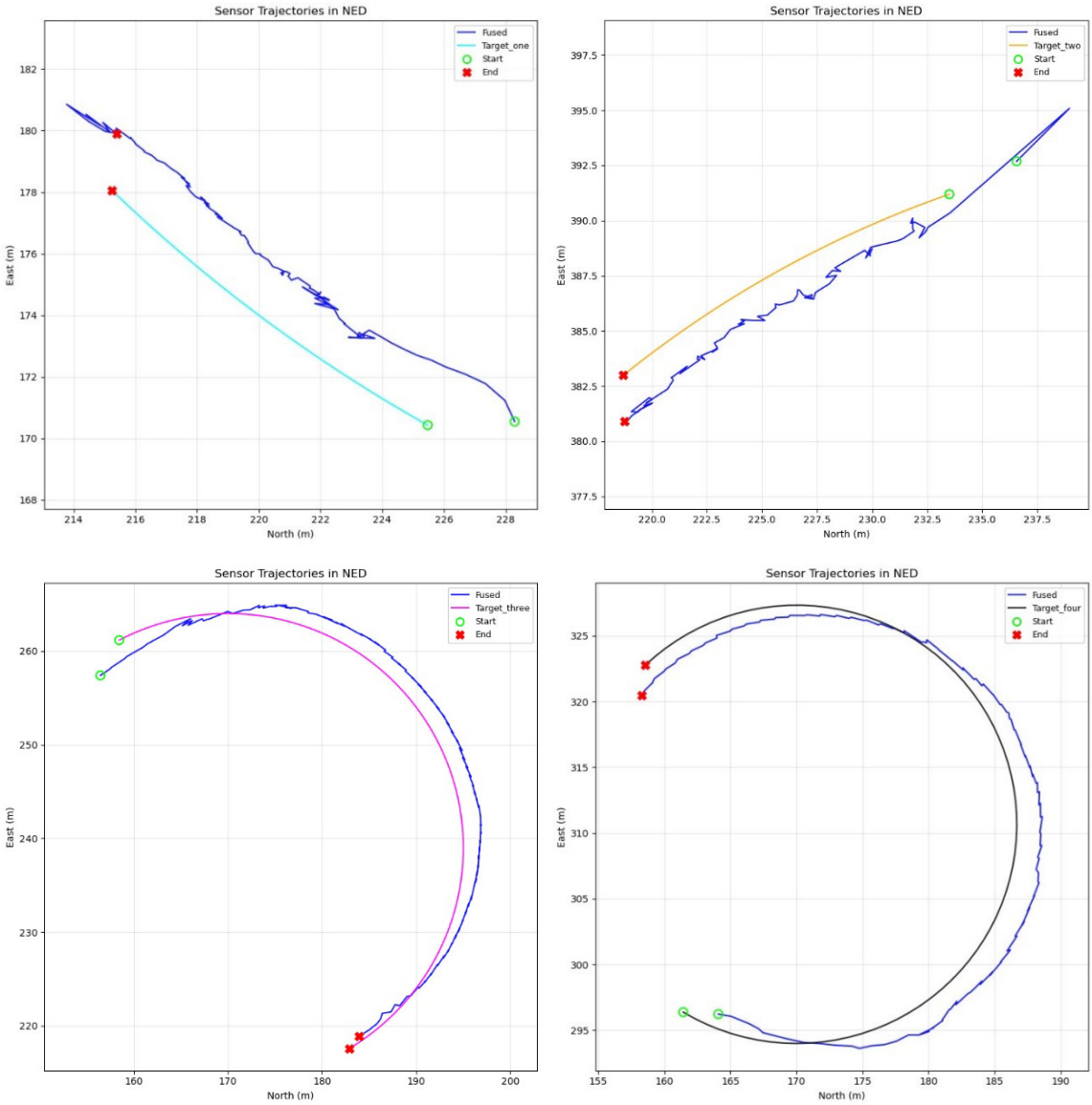


Figure 36 - Ground truth trajectories of the four vessel targets – target one (upper left), target two (upper right), target three (lower left) and target four (lower right) – vs. the Kalman filter output for straight path trajectory in NED coordinates.

As observed, the estimated trajectories follow the ground truth closely across the entire sequence, with no evident track divergence. Minor oscillations appear when the WIG approaches or leaves the target cluster, where detection geometry becomes less favourable. Despite this, the filter maintains stable updates and properly distinguishes between nearby targets, demonstrating effective multi-target tracking.

The corresponding distance errors for the same trajectory are shown in Figure 37. Across all targets, the average positional error remains between 2 and 3 meters, consistent with the expected centroid offset introduced by the vessel size and viewing perspective.

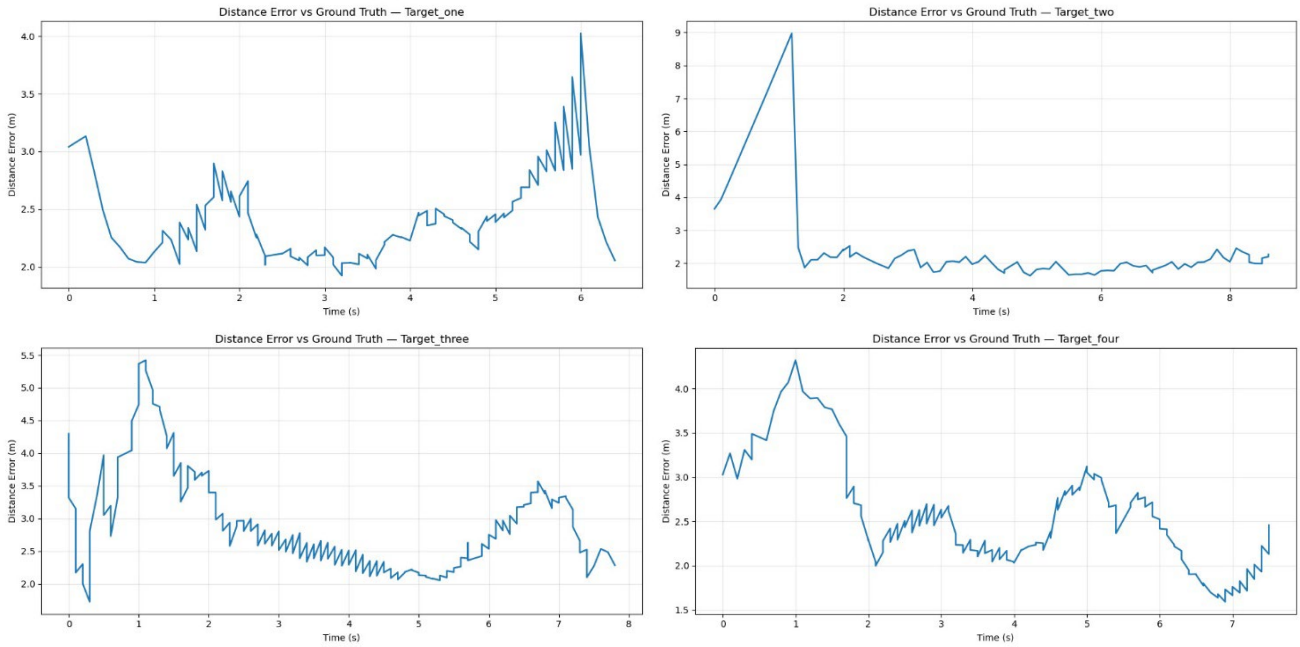


Figure 37 - Distance errors between ground truth and Kalman filter estimates for each vessel during the straight path trajectory

These results indicate that under steady flight conditions and nearly linear motion, the Kalman filter achieves high positional accuracy and stable convergence, even when detections are partially affected by centroid variation.

The orbital trajectory scenario imposed more challenging dynamics, as the WIG continuously changed orientation while circling around the target area. The same four vessels followed the same circular motion patterns described earlier. The corresponding tracking results are shown in Figure 38, comparing ground truth and filter estimates.

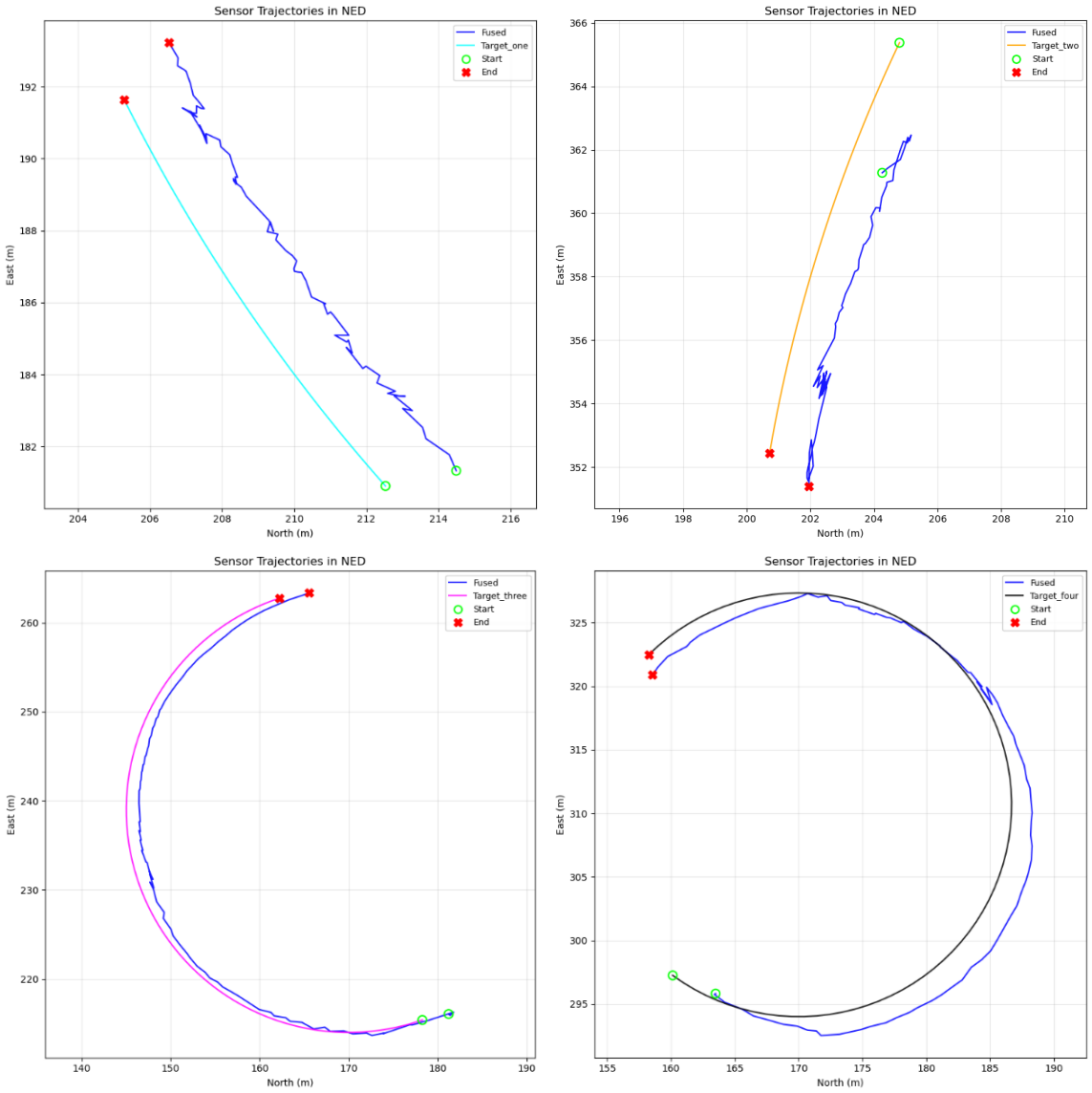


Figure 38 - Ground truth trajectories of the four vessel targets – target one (upper left), target two (upper right), target three (lower left) and target four (lower right) – vs. the Kalman filter output for orbital trajectory in NED coordinates.

Despite more complex relative motion and greater viewpoint variation, the Kalman filter successfully preserved target association and continuity. Small deviations are observed when the WIG faces away from the targets or when the vessels become temporarily aligned in bearing, but the estimated trajectories rapidly, re-converge once better sensor geometry is restored.

The distance error plots for the orbital trajectory are provided in Figure 39. Similar to the straight path results, the position error remains within 2-3 meters for most of the simulation. Slightly higher peaks appear occasionally, corresponding to short periods of reduced target visibility or bearing overlap.

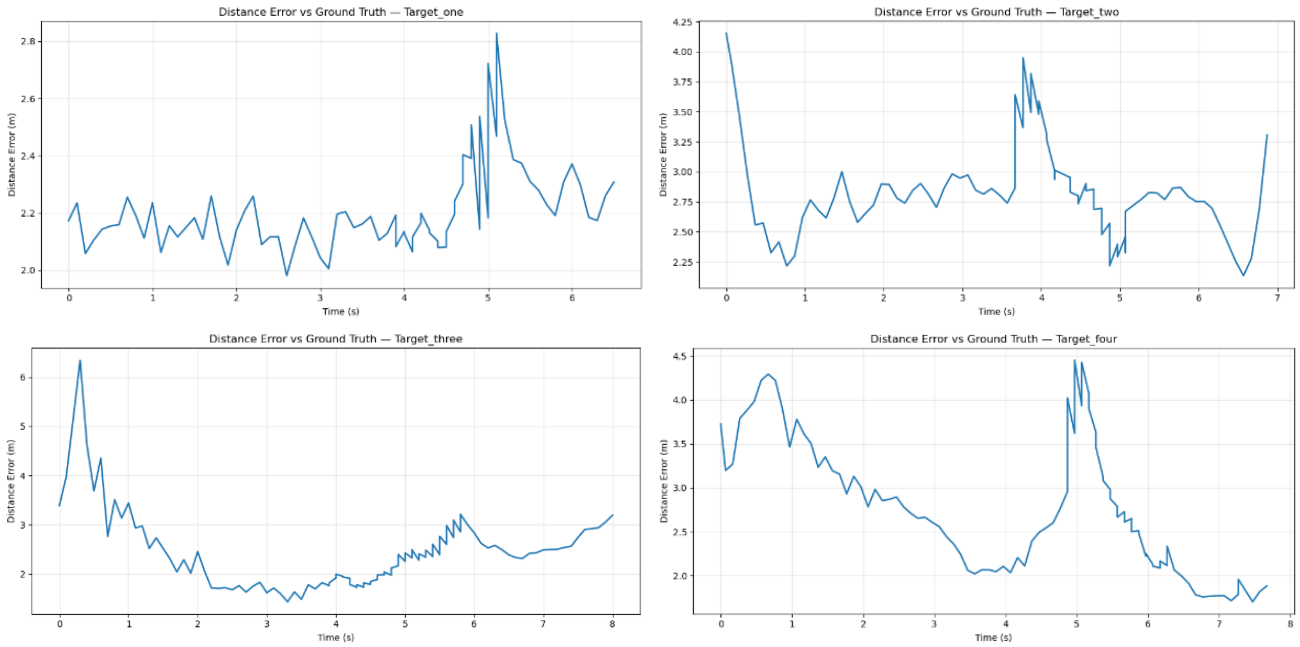


Figure 39 - Distance errors between ground truth and Kalman filter estimates for each vessel during the orbital trajectory

Overall, these findings confirm that the Kalman filter is capable of maintaining accurate and stable multi-target tracking across varying motion profiles, with the observed errors well aligned with the geometric and detection uncertainties inherent to the simulated setup.

8 Surface Modelling

The primary operational environment of an AIRSHIP's WIG vehicle is the water surface. Consequently, an accurate and dynamic representation of this surface is essential for safe and efficient flight. The altitude at which a WIG operates is intrinsically linked to the ground-effect region, where aerodynamic benefits such as reduced induced drag significantly lower power consumption. To exploit these advantages, the vehicle must maintain flight at minimal but safe clearance above the water.

However, the sea surface is not static. Its shape varies continuously as a function of meteorological and oceanographic conditions, including wind intensity, swell direction, wave spectra, and local disturbances. These variations introduce surface undulations that define the physical lower boundary of the flight envelope. For a WIG following a straight path, insufficient knowledge of the instantaneous or forecasted water surface elevation may lead to loss of aerodynamic efficiency due to excessive altitude, or hazardous proximity to wave crests, potentially resulting in unintended surface contact.

Therefore, surface modelling is a critical task that enables the prediction, estimation, and reconstruction of the water surface profile along the planned trajectory. Such a model may support real-time altitude control, trajectory planning, and operational safety.

Task T6.3 is still ongoing, and the work performed so far reflects the early stages of establishing a reliable water-surface representation for WIG operations. The progress achieved to date will be described in the following subsections, along with a brief outline of the next steps required to further advance the modelling framework.

8.1 Wave Detection Technologies

Many sectors, such as offshore energy, maritime transport, and coastal monitoring, require accurate wave detection to support safe operations and environmental assessments. These industries have therefore developed a wide range of non-contact sensing technologies that may be used for WIG operations.

The current approaches for the contactless wave perception include RADARs, LiDARs and vision-based technologies (Zhou, Liu, Wu, & Guo, 2025).

Marine X-band radars are widely installed on ships and offshore platforms, providing large-area wave-field information by analysing backscattered radar images. They operate reliably in most weather conditions, but require complex inversion algorithms to extract wave elevation, and their accuracy depends strongly on the sea state and proper calibration. Downward-looking LiDAR systems offer very high vertical accuracy by measuring the distance to the water surface using laser pulses. They are effective for profiling, but their performance deteriorates under spray, fog, or rain, and their coverage is limited unless mounted on a scanning mechanism. Vision-based systems, including standard cameras, stereo imaging, and infrared sensors, reconstruct the sea-surface shape through image processing. They are inexpensive and provide dense spatial information, but their reliability is highly sensitive to illumination, reflections, and atmospheric conditions, and they require significant computational resources for real-time operation.

Overall, no single sensing method is universally optimal. Each technology presents trade-offs between accuracy, spatial coverage, environmental sensitivity, and computational demands. Radar systems offer robustness and wide coverage but rely on heavy processing and assumptions in the inversion model. LiDAR provides very precise point measurements but is limited by adverse weather conditions and a narrow field of view. Vision-based systems deliver detailed surface information but are the most dependent on environmental conditions and require extensive computation. As a result, the use of a single modality to detect waves is insufficient, and the best approach is to fuse data from multiple sensors.

8.2 Linear Wave Theory

Given the limitations of the WIG perception system described in the previous section, the onboard sensors alone cannot provide a complete and continuous reconstruction of the water surface. The camera offers valuable but indirect information about wave shape, and the altimeter provides only a single range measurement. As a result, the perception system requires a mathematical model capable of interpreting sparse or partial observations and producing a coherent estimate of the sea surface along the WIG's flight path.

To address this, the adoption of Linear Wave Theory as the basis for water-surface modelling is required. This theory offers a straightforward yet physically meaningful description of wave motion and is widely employed in ocean engineering due to its analytical clarity and computational efficiency (Klein, et al., 2020).

In its basic form, Linear Wave Theory represents the water-surface elevation $\eta(x,t)$ as a sinusoidal function:

$$\eta(x,t) = a\cos(kx - \omega t + \varphi)$$

Where a is the wave amplitude, k is the wave number, ω is the angular frequency, and φ is the phase. his equation describes a travelling wave whose shape, speed, and height depend on measurable parameters that can be estimated from the WIG sensors. Although actual sea states are composed of many wave components

interacting simultaneously, Linear Wave Theory remains relevant because it allows the surface to be approximated as the sum of multiple linear waves:

$$\eta(x,t) = \sum_i a_i \cos(k_i x - \omega_i t + \varphi_i)$$

This superposition principle enables the system to represent a wide variety of sea states using a relatively small number of wave components, each corresponding to a dominant frequency observed in sensor data.

The meaning of the model is straightforward: the wave amplitude determines the maximum elevation of the water surface relative to mean sea level, the wave number sets the wavelength, and the angular frequency governs the rate at which the wave pattern moves forward. By adjusting these parameters, the model can reproduce the essential characteristics of the surface undulations that affect the WIG's safe operating altitude.

While Linear Wave Theory cannot capture highly nonlinear wave phenomena such as breaking waves, steep crests, or strong turbulence, it provides a sufficiently accurate representation for moderate sea states, which are the conditions most relevant for typical WIG operations. Moreover, its simplicity allows real-time implementation and makes it well-suited for fusion with the limited but complementary information available from the WIG sensors. In this way, the model serves as the backbone for estimating the surface profile ahead of the vehicle and for predicting the minimum safe altitude during ground-effect flight.

8.3 WIG Sensors for Wave Perception

The WIG perception system utilizes a diverse set of sensors, enabling the reliable detection of the environment under various weather and lighting conditions. This suggests that the system has capabilities similar to those described in Section 8.1, such as marine X-band RADAR or scanning LiDAR profilers. However, the WIG platform cannot directly duplicate these solutions or use them because they are either physically large or unsuitable for operation on a fast, low-altitude aircraft. Therefore, only some technologies commonly used at sea have practical equivalents in the WIG's onboard perception system.

The RADAR integrated into the WIG system is a millimetric-wave RADAR, originally intended for obstacle detection in front of the vehicle. This type of RADAR is compact, lightweight, and capable of high-resolution ranging at short distances, making it more suitable for a fast-moving platform compared to marine radars. Marine RADAR, although effective for large-area wave monitoring on ships, has slower scan rates, larger antennas, and processing pipelines optimized for stationary or slow-moving vessels. Their characteristics are therefore incompatible with the dynamic and real-time requirements of a WIG, making millimetric RADAR the only practical choice, even though it does not provide the broad surface-state information that marine RADAR can deliver.

A similar limitation exists for LiDAR. The main LiDAR installed on the WIG is oriented and configured for target detection rather than for nadir-looking surface profiling. Its beams are not perpendicular to the water surface, meaning that almost no useful returns are obtained from the sea except in the presence of foam and some wave crests, where reflectivity is higher.

Vision-based sensing is the primary option for reconstructing the sea surface ahead of the vehicle. Although vision is not directly comparable to the specialized multi-beam or scanning LiDAR systems used in marine wave measurements, it can still provide valuable and rich information on the geometry and motion of the water surface when processed appropriately. In addition to the cameras, the WIG perception system includes an altimeter capable of measuring the distance to the water surface. Although this measurement is limited to a single beam, it can be combined with the image data to enhance the estimation of local wave elevation.

Together, these sensors form a perception framework that, while more constrained than those used in maritime applications, still provides a viable basis for surface-state reconstruction and wave prediction during low-altitude flight.

8.4 Current Results

To date, the primary focus has been on reconstructing the three-dimensional structure of the sea surface from the output of the EO cameras acquired by the stereo camera pair. Since vision-based perception is the primary method available on the WIG for observing the water surface, obtaining an accurate and consistent 3D reconstruction is an essential step before integrating any wave-modelling approach.

The first part of the work concentrated on the precise calibration of the stereo cameras. This included determining both the intrinsic parameters of each camera, such as focal length, principal point, and lens distortion, as well as the extrinsic parameters that describe the relative position and orientation between the two sensors. Accurate calibration is critical because even small errors in these parameters can introduce significant inaccuracies in the depth estimation, especially when reconstructing surfaces with low texture, such as open water.

Following the intrinsic and extrinsic calibration, effort was dedicated to calibrating the disparity map computation. This process involved tuning the stereo-matching algorithm to the specific characteristics of the WIG camera setup and to the visual properties of maritime scenes. Challenges such as low contrast, specular reflections, and the dynamic nature of the water surface required careful adjustment of the matching parameters to obtain stable and consistent disparity estimations. Several configurations were tested to reduce noise, improve depth continuity, and minimize false matches in regions where the water texture provides limited feature information.

The experiments have demonstrated that, although the sea surface is a difficult target for stereo reconstruction, it is still possible to retrieve useful depth information when conditions allow for sufficient visual structure. The combination of refined calibration, optimized stereo processing, and tailored filtering techniques has yielded preliminary 3D models that accurately capture the general shape of the observed surface, see Figure 40.



Figure 40 - Sea wave reconstruction from stereo vision.

These results form a solid foundation for the next phase of the work, in which the reconstructed point clouds will serve as a key input to an overall wave-modelling framework. Given the direction of motion of the vehicle, wave crests can be extracted from the point-cloud data, enabling the definition of a safe flight altitude for the WIG vehicle. However, to achieve more accurate estimation and reconstruction, higher-rate measurements are required. Considering the sensor suite available on the WIG platform, the altimeter is particularly well-suited for this purpose. Its high sampling rate enables the acquisition of multiple echoes from the water surface, which can

be used to enhance the stereo-based point cloud and improve the forecasting of incoming waves through the modelling of previously observed wave patterns.

In addition to active ranging sensors, passive sensing should not be overlooked. Thermal imaging, for instance, has the potential to reveal wave structures under suitable environmental conditions, as temperature differences between wave crests, troughs, and surrounding water may provide complementary visual cues. Such information could support the identification of dominant wave features, particularly in scenarios where geometric reconstruction alone is challenging.

Finally, it is worth noting that the current LiDAR configuration is expected to return valid measurements primarily in the presence of surface foam, which typically occurs near wave crests where reflectivity is higher. While this limits the spatial completeness of the LiDAR data, it may nonetheless provide valuable and reliable indications of critical wave features, reinforcing the detection of crests that are most relevant for safe low-altitude flight.

9 Conclusion

This report addressed the challenge of providing reliable, real-time perception capabilities for autonomous WIG vehicles operating in complex and dynamic maritime environments. Safe WIG operation requires low-latency, long-range awareness of obstacles and accurate modelling of the surrounding environment, including the sea surface, to enable effective collision avoidance and stable ground-effect flight.

Within WP6, a suitable suite of heterogeneous sensors was selected, integrated, and refined to meet these requirements. As described in this deliverable, RADAR, LiDAR, optical and thermal cameras, together with high-rate navigation sensors, were processed and combined through a real-time multimodal sensor fusion framework. This framework supports both robust localization and reliable obstacle detection and tracking, while guaranteeing operation within the real-time constraints imposed by autonomous WIG flight.

The developed perception system was extensively tested and evaluated. Real-world experiments were conducted to validate system performance under realistic operating conditions, while simulation-based testing was employed for scenarios that were impractical to reproduce experimentally. These complementary evaluation approaches ensured broad coverage of operational conditions and failure cases.

Overall, the results demonstrate that the proposed perception architecture performs in accordance with the project plan and the defined technical requirements, providing a solid base for the collision avoidance functions developed in WP7 and for future autonomous WIG operations. Final tuning and adaptation of the system are expected to take place during integrated operation within the WIG platform, where vehicle-specific dynamics and mission profiles can be fully accounted for.

References

Almeida, C. (2009). *Sistema de percepção visual para veículos autónomos aéreos.* Instituto Politecnico do Porto (Portugal).

Bar-Shalom, Y., Daum, F., & Huang, J. (2009). The probabilistic data association filter. *IEEE Control Systems Magazine, 29*, 82-100. doi:10.1109/MCS.2009.934469

Bovcon, B., & Kristan, M. (2021). WaSR—A water segmentation and refinement maritime obstacle detection network. *IEEE Transactions on Cybernetics, 52*, 12661--12674.

Chen, X., Liu, Y., & Achuthan, K. (2021). WODIS: Water obstacle detection network based on image segmentation for autonomous surface vehicles in maritime environments. *IEEE Transactions on Instrumentation and Measurement, 70*, 1--13.

Cho, H., Seo, Y.-W., Kumar, B., & Rajkumar, R. (2014). *A multi-sensor fusion system for moving object detection and tracking in urban driving environments.* doi:10.1109/ICRA.2014.6907100

Cooper, C., Zhang, J., Ragai, I., & Gao, R. (2024). Multi-sensor fusion and machine learning-driven sequence-to-sequence translation for interpretable process signature prediction in machining. *Journal of Manufacturing Systems*, 288--298. doi:10.1016/j.jmsy.2024.04.010

Helgesen, Ø., Brekke, E., Helgesen, H., & Engelhardtsen, Ø. (2019). Sensor Combinations in Heterogeneous Multi-sensor Fusion for Maritime Target Tracking. In *2019 22th International Conference on Information Fusion (FUSION)* (pp. 1-9). doi:10.23919/FUSION43075.2019.9011297

Hu, J.-w., Zheng, B.-y., Wang, C., Zhao, C.-h., Hou, X.-l., Pan, Q., & Xu, Z. (2020). A survey on multi-sensor fusion based obstacle detection for intelligent ground vehicles in off-road environments. *Frontiers of Information Technology & Electronic Engineering, 21*, 675--692. doi:10.1631/fitee.1900518

Klein, M., Dudek, M., Clauss, G., Ehlers, S., Behrendt, J., Hoffmann, N., & Onorato, M. (2020). On the deterministic prediction of water waves. *Fluids, 5*, 9.

Rodríguez Flórez, S., Frémont, V., Bonnifait, P., & Cherfaoui, V. (2014). Multi-modal object detection and localization for high integrity driving assistance. *Machine Vision and Applications, 25*, 583--598.

Zhou, Z., Liu, J., Wu, R., & Guo, A. (2025). Non-contact sensing methods for wave measurement: current research and challenge. *Intelligent Transportation Infrastructure*.