Class-Agnostic Maritime Obstacle Detection with Simulated IMU-Guided Horizon Stabilization for Autonomous Surface Vessels

Ananda Sangli
Khoury College of Computer Sciences
Northeastern University
Boston, USA
sangli.a@northeastern.edu

Aryan Alpeshkumar Patel
Khoury College of Computer Sciences
Northeastern University
Boston, USA
patel.aryana@northeastern.edu

Julee Chung
Khoury College of Computer Sciences
Northeastern University
Boston, USA
chung.woo@northeastern.edu

***Abstract*—Autonomous Surface Vessels (ASVs) must detect an open-ended set of hazards from a camera whose viewpoint shifts continuously with vessel motion, as wave-induced pitch and roll destabilize the horizon and displace small, distant obstacles between frames. We take a class-agnostic approach that collapses all hazards into a single "obstacle" class and build a unified benchmark from the USVTrack and LaRS datasets. A YOLOv8n detector performs strongly within-domain on USVTrack but worse on LaRS, exposing a large cross-domain generalization gap. Because the datasets provide no synchronized per-frame IMU telemetry, we evaluate camera-motion compensation with a controlled synthetic-perturbation framework: frames are subjected to a known, time-correlated pitch/roll signal, while stabilization is driven only by a simulated noisy IMU estimate rather than the ground-truth angle. Using a frozen detector, motion at $\pm 30^\circ$ roll / $\pm 15^\circ$ pitch reduces recall from 0.891 to 0.561, while stabilization increases recall to 0.723; the benefit grows with motion severity. A constant-velocity Kalman filter further reduces bounding-box jitter by 32% on the unstabilized stream and 43% on the stabilized stream, while improving identity stability without detection recall. We note the reliance on synthetic motion as a limitation of this study and outline the use of real IMU data and real-time deployment as future work.**

***Index Terms*—maritime perception, obstacle detection and tracking, autonomous surface vessels, horizon stabilization, inertial measurement unit, Kalman filter, YOLOv8**

I. Introduction

Autonomous Surface Vessels (ASVs) operating on lakes, rivers, and coastal waters must perceive hazards from a camera mounted on a platform that is continuously affected by vessel motion. Wave-induced pitching and rolling, spray, glare, and changing viewpoints can make visual perception substantially more difficult than in conventional road-scene environments. In particular, three properties distinguish maritime perception from typical autonomous-driving scenarios. First, the horizon is unstable, causing the apparent image position of an object to change between frames even when the underlying scene is static. Second, maritime hazards are open-ended and may include buoys, debris, swimmers, moored vessels, and moving vessels, making a fixed semantic taxonomy less appropriate for collision avoidance. Third, small and distant obstacles are particularly safety-critical but are often detected with low confidence and can easily be missed.

To address the open-ended nature of maritime hazards, we adopt a class-agnostic detection strategy in which all retained annotated objects are mapped to a single "obstacle" class. This formulation removes the need to distinguish between specific semantic categories and instead focuses the detector on the operational question of whether an obstacle is present.

Our work investigates two complementary mechanisms for improving perception under maritime motion. The first is horizon stabilization, which attempts to compensate for camera pitch and roll before object detection. The second is temporal state estimation, which uses a Kalman filter to smooth object trajectories and maintain tracks through short detection gaps. We formulate these as two research questions:

- **RQ1 Horizon Stabilization:** Does compensating for camera pitch and roll improve accuracy under maritime camera motion?
- **RQ2 State Estimation:** Does a Kalman filter reduce bounding-box jitter and improve track continuity following intermittent missed detections?

A key practical limitation is that the datasets used in this project do not expose synchronized per-frame pitch and roll IMU measurements. Rather than treating the lack of telemetry as a blocker, we developed a modular synthetic-motion pipeline. The pipeline first applies a controlled camera-motion perturbation to an input frame and then generates a corresponding synthetic IMU-derived orientation estimate containing sensor-like bias, noise, and latency. The stabilizer receives this imperfect estimate rather than the ground-truth transformation used to generate the perturbed image. This provides a controlled way to evaluate the motion-compensation mechanism without directly supplying the correction that generated the perturbation.

The resulting system separates the perception problem into three stages: detection, motion compensation, and temporal tracking. A frozen YOLOv8n detector provides the same detection model for all downstream experiments, allowing the effects of stabilization and tracking to be evaluated independently. The stabilization experiment measures whether correcting camera motion recovers detection performance, while the tracking experiment evaluates whether temporal filtering produces smoother and more stable trajectories.

II. Related Work

Maritime obstacle detection and datasets. Our dataset selection follows recent work on maritime perception datasets and technologies for autonomous surface vessels [1]. We use two complementary benchmarks: USVTrack [2], which provides sequential data suitable for tracking-oriented evaluation in inland and coastal environments, and LaRS [3], a diverse panoptic maritime obstacle dataset spanning lakes, rivers, and coastal scenes.

USVTrack is particularly useful for the temporal experiments because it provides sequential frames and tracking information. LaRS contributes additional visual diversity through different environments and maritime backgrounds. Combining the two datasets allows the detector to be trained on a broader range of maritime appearances while retaining USVTrack sequences for controlled temporal evaluation.

IMU-Assisted Horizon Stabilization. Prior work has investigated the use of inertial measurements to compensate for camera motion in maritime perception. Bovcon et al. [4], for example, incorporate IMU pitch and roll information into semantic segmentation to estimate the water horizon under vessel motion. WaSR [5] similarly incorporates maritime-specific visual information for water and obstacle segmentation.

Our approach differs in several respects. First, we use a class-agnostic object detector rather than a semantic segmentation model. Second, motion compensation is performed as a geometric preprocessing operation before neural-network inference rather than being fused into the network architecture. Finally, because the datasets used here do not expose synchronized per-frame pitch and roll measurements, we use a controlled synthetic-motion framework with an imperfect synthetic IMU-derived orientation estimate.

Detection and Tracking. Our detector is based on YOLOv8 [6], which provides a lightweight object-detection architecture suitable for eventual real-time deployment. For tracking, we follow the general structure of SORT [7], which combines a constant-velocity Kalman filter [8] with IoU-based association.

Rather than using a Kalman filter as part of the detector itself, we apply it after detection. This allows the experiment to separately evaluate whether temporal filtering improves trajectory smoothness and track continuity without changing the underlying detector.

III. Method

A. System Overview

The proposed perception pipeline combines a class-agnostic YOLOv8n detector with two downstream modules: a pre-inference horizon stabilizer for RQ1 and a Kalman-based tracker for RQ2. The same frozen detector is used throughout the experiments, while stabilization and tracking are independently enabled or disabled.

The overall experimental pipeline is:

Input maritime frame $\rightarrow$ simulated camera motion $\rightarrow$ optional horizon stabilization $\rightarrow$ YOLOv8n detection $\rightarrow$ optional Kalman tracking.

This design allows the effects of each module to be isolated. In particular, comparing unstabilized and stabilized frames measures the contribution of horizon compensation, while comparing filter-free and Kalman-based tracking measures the contribution of temporal state estimation.

B. Class-Agnostic Obstacle Detection

All retained annotations are converted into a single category, `Class 0: obstacle`. Obstacle-free frames are represented by empty label files rather than by an explicit background class. A single YOLOv8n detector is trained on the resulting unified dataset.

The class-agnostic formulation resolves the label mismatch between the two source datasets and focuses the detector on obstacle presence rather than semantic identity. This is particularly appropriate for the intended collision-avoidance application, where an unfamiliar or previously unseen object may still represent a navigational hazard.

C. Unified Single-Class Benchmark

The training combines USVTrack and LaRS. USVTrack provides sequential frames and tracking ground truth, making it the primary temporal domain for the stabilization and tracking experiments. LaRS contributes additional visual diversity through varied maritime environments and obstacle appearances.

The original datasets are kept read-only, while a generated unified corpus serves as the detector's training input. For USVTrack, tracking annotations are parsed directly, and timestamp-named images are paired with frame indices positionally. For LaRS, bounding boxes are extracted from the panoptic annotations for the retained dynamic "thing" classes, while "stuff" categories are discarded.

The shoreline-spanning *static obstacle* category in LaRS is excluded because its bounding box covers the waterline and does not correspond well to the object-level obstacle formulation used in this project. As a result, the trained detector focuses on dynamic obstacles.

USVTrack is resampled at a stride of five frames. Five complete sequences are held out for validation to reduce temporal leakage between training and validation. Because the validation set is USVTrack-heavy, detector performance is also reported separately by dataset when evaluating cross-domain generalization.

D. Synthetic IMU Perturbation and Horizon Stabilization

The available datasets do not provide synchronized per-frame IMU measurements. Consequently, direct evaluation using recorded vessel telemetry was not possible. We therefore developed a modular synthetic-motion pipeline to validate the stabilization approach under controlled camera motion.

The pipeline separates the generation of camera motion from the estimation of that motion. A camera-roll simulator begins with an undistorted maritime frame and applies a time-varying rotational perturbation designed to emulate wave-induced camera motion. The perturbation is generated independently from the stabilization stage, allowing both the magnitude and temporal behavior of the simulated motion to be controlled.

A corresponding synthetic IMU-derived orientation measurement is then generated from the simulated motion. To approximate sensor imperfections, the measurement includes a fixed bias, Gaussian measurement noise, and a one-frame temporal latency. Importantly, the stabilization stage does not receive the ground-truth rotation used to generate the image. It instead receives the imperfect synthetic measurement.

The stabilizer converts the estimated camera orientation into an image transformation and applies the corresponding inverse transformation to the perturbed frame. The corrected image is then passed to the frozen YOLOv8n detector.

This separation provides two complementary benefits. First, the camera-motion simulator gives the experiment known ground truth, allowing the amount of perturbation to be controlled. Second, the stabilizer must operate using only an imperfect motion estimate, preventing the evaluation from trivially providing the exact correction used to generate the perturbed image.

The synthetic IMU is therefore not intended to reproduce every characteristic of a physical inertial sensor. Instead, it provides a controlled intermediate validation of the geometric motion-compensation pipeline. Real IMU telemetry remains necessary for evaluating effects such as sensor drift, vibration, frame synchronization, and camera-to-IMU alignment.

E. Controlled Camera Motion Evaluation

To quantify the effect of camera motion on detection, we evaluate the frozen detector under controlled, time-varying pitch and roll perturbations. The camera motion is represented using sinusoidal signals, with the resulting perturbation applied as a rotational homography based on the camera intrinsic matrix.

The same perturbed frames are used across all stabilization ablations so that changes in performance are attributable to the stabilization and tracking modules rather than differences in the input sequence.

Five experimental configurations are evaluated:

1) **Clean:** Unperturbed input with no stabilization or tracking.
2) **A – Baseline:** Perturbed input with neither stabilization nor Kalman filtering.
3) **B – RQ1 Stabilization:** Perturbed input corrected using the synthetic IMU-derived orientation estimate before detection.
4) **C – RQ2 Kalman:** Perturbed input without stabilization but with Kalman tracking.
5) **D – Combined:** Perturbed input with both stabilization and Kalman tracking.

The clean condition establishes the performance ceiling on the original frames. Configuration A measures the degradation caused by camera motion. Configuration B isolates the contribution of horizon stabilization, while Configurations C and D evaluate the effect of temporal state estimation with and without stabilization.

Ground-truth bounding boxes are transformed using the same image transformation applied to the corresponding frames, ensuring that evaluation remains aligned with the perturbed image geometry.

F. Object Tracking: Filter-Free vs. Kalman

Two tracking configurations are evaluated using the same detector outputs. The filter-free baseline performs greedy Intersection-over-Union (IoU) association between consecutive detections without a motion model or coasting. Consequently, a missed detection terminates the existing track, and a subsequent detection may be assigned a new identity. The Kalman tracker maintains a constant-velocity state consisting of the bounding-box center coordinates, width, height, and center velocities. The state vector is defined as

$$\mathbf{x}_t = \begin{bmatrix} c_x & c_y & w & h & v_x & v_y \end{bmatrix}^\top, \tag{1}$$

where c_x and c_y denote the bounding-box center coordinates, w and h denote its width and height, and v_x and v_y represent the corresponding center velocities. The predicted state is associated with new detections using greedy IoU matching. When a detection is temporarily unavailable, the tracker can coast using the predicted state for a bounded number of frames before the track is deleted. Both trackers operate exclusively on bounding-box geometry. This design isolates the contribution of temporal state estimation from any appearance-based re-identification mechanism.

IV. Experimental Setup

A. Datasets and Metrics

The unified benchmark contains 8,140 training frames and 1,535 validation frames after class consolidation and USVTrack resampling. The training split contains 5,535 USVTrack frames and 2,605 LaRS frames, while the validation split contains 1,337 USVTrack frames and 198 LaRS frames.

TABLE I
Unified single-class benchmark: frame counts per split and source.

Split	USVTrack	LaRS	Total
Train	5,535	2,605	8,140
Val	1,337	198	1,535

Detector quality is evaluated using precision, recall, mean average precision at IoU 0.5 (mAP@50), and mean average precision over IoU thresholds from 0.5 to 0.95 (mAP@50–95).

For the stabilization and tracking experiments, recall at IoU 0.5 is emphasized because the primary question is whether obstacles lost under camera motion can be recovered by stabilization or temporal processing.

Temporal consistency for RQ1 is evaluated using the per-sequence variation in recall. For RQ2, bounding-box jitter is measured using the mean frame-to-frame centroid acceleration of a track. Dropout recovery is measured as the fraction of detector misses that are bridged by a coasted Kalman track.

These stabilization and tracking metrics are computed using the project's fixed-confidence evaluation harness rather than the Ultralytics validation procedure. This keeps the custom temporal metrics separate from the standard detector metrics.

B. Implementation Details

The detector is YOLOv8n trained for 100 epochs at 640×640 with batch size 32. Training was configured for up to 100 epochs with early stopping using a patience of 20 epochs. Training converged around epoch 48, and early stopping terminated training at epoch 68. The resulting detector was exported to ONNX with opset 12 and frozen for all downstream experiments.

The synthetic camera perturbation uses time-correlated pitch and roll signals composed of sinusoidal components at 0.15 Hz and 0.25 Hz. The perturbation is represented as a rotational homography under an assumed horizontal field of view of $80°$. Unless otherwise specified, the primary stabilization ablation uses a roll amplitude of $30°$ and pitch amplitude of $15°$.

The synthetic IMU-derived orientation estimate is generated by adding a $0.5°$ bias, $1.0°$ Gaussian measurement noise, and one frame of latency to the simulated orientation. The Kalman tracker uses a constant-velocity model with greedy IoU association at a threshold of 0.30, a maximum coast age of one frame, and a minimum of two hits required to confirm a track.

All downstream experiment scripts run inference through ONNX Runtime on CPU, and random seeds are fixed for reproducibility.

V. Results

A. Overall Detection Performance

On the combined validation set, the trained detector achieves an mAP@50 of 0.756 and an mAP@50–95 of 0.440, with precision of 0.860 and recall of 0.700 (Table II).

The detector performs substantially better within the USVTrack domain than on the LaRS domain, demonstrating a cross-domain generalization gap. This difference motivates reporting results by source dataset rather than relying exclusively on a single blended validation score.

The normalized confusion matrix indicates that approximately 85% of obstacles are correctly detected, while approximately 15% are missed as background. These results show that localization and precision are relatively strong, while recall remains the primary weakness. This is particularly relevant for collision avoidance because small, distant, or partially submerged obstacles can be difficult to detect, and missed obstacles are more consequential than additional false positives.

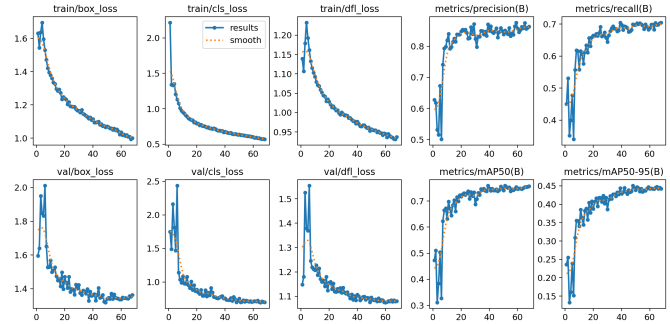

Fig. 1. Training and validation curves over 68 epochs. Box, classification, and DFL losses decrease steadily on both splits (validation losses spike briefly in the first few epochs before settling), while validation precision, recall, mAP@50, and mAP@50–95 rise and plateau. Training was early-stopped after 20 epochs without improvement; the best checkpoint is from epoch 48. Metrics are computed on the combined validation set.

Training converged without overfitting: the validation losses plateaued without divergence and the detection metrics flattened by roughly epoch 48, before early stopping at epoch 68.

B. Qualitative Detections

Fig. 2 shows the model's predictions on held-out USVTrack frames. Qualitative evaluation on held-out USVTrack frames shows that the detector can localize small and distant vessels against complex urban shorelines while assigning the unified "obstacle" class (roughly 0.4-0.8 confidence).

C. Confidence-Threshold Operating Point

We evaluate confidence thresholds of 0.15, 0.25, and 0.35 to select the operating point for downstream experiments. A lower threshold is considered because missed obstacles are more costly than additional false positives for the intended maritime collision-avoidance application.

The number of detected obstacles increases as the threshold decreases. Qualitatively, a threshold of 0.15 produces more persistent detections across consecutive frames, while thresholds of 0.25 and 0.35 cause some smaller and more distant objects to appear intermittently.

The F1-confidence curve reaches a maximum of approximately 0.77 near a confidence threshold of 0.34, but remains relatively broad and flat over a substantial confidence range. We therefore select $\tau_{\mathrm{conf}} = 0.15$ for the stabilization comparisons. This operating point sacrifices little overall F1 while favoring recall for the small, distant targets that are most likely to be missed.

D. RQ1: Effect of Horizon Stabilization

We evaluate horizon stabilization on the 1,337 USVTrack validation frames using four configurations with a common frozen detector. The clean condition provides an unperturbed reference, while configurations A–D use the same controlled camera-motion perturbations.

On unperturbed frames, the detector achieves a recall of 0.891, establishing the reference ceiling. Under the primary perturbation of $\pm 30°$ roll and $\pm 15°$ pitch, recall decreases to 0.561 without stabilization. This represents a reduction of 0.330 caused by the simulated camera motion.

Adding horizon stabilization increases recall from 0.561 to 0.723, recovering 0.162 of the lost recall, or approximately half of the motion-induced degradation. Stabilization also improves precision relative to the unstabilized motion condition.

The Kalman-only configuration produces a recall of 0.566, which is essentially unchanged from the unstabilized baseline. This is expected because the Kalman filter operates after detection and therefore cannot recover an obstacle that the detector failed to observe in the first place. The combined configuration reaches 0.726, only slightly above stabilization alone at 0.723. Thus, stabilization is responsible for the principal recall improvement, while Kalman filtering primarily affects the temporal quality of the detections.

The motion-amplitude sweep further demonstrates that the stabilization benefit depends on motion severity. As roll amplitude increases from $\pm 5°$ to $\pm 30°$, the recall gap between the unstabilized and stabilized configurations increases from approximately 0.006 to 0.162. The stabilized configuration remains close to the clean reference, while the unstabilized baseline degrades increasingly with motion. This indicates that horizon stabilization provides its greatest benefit when camera motion causes substantial detector degradation.

TABLE II
FOUR-CONFIGURATION ABLATION UNDER MOTION (ROLL $\pm 30°$, PITCH $\pm 15°$), USVTRACK VALIDATION. CLEAN IS THE UNPERTURBED REFERENCE. STABILIZATION (B) RECOVERS ABOUT HALF THE MOTION-INDUCED RECALL LOSS; KALMAN (C) LEAVES RECALL ESSENTIALLY UNCHANGED.

Config	Stab.	Kalman	R	P	F1	mIoU
CLEAN (ceiling)	–	–	0.891	0.875	0.882	0.815
A (baseline)	no	no	0.561	0.763	0.646	0.746
B (RQ1)	yes	no	0.723	0.833	0.774	0.750
C (RQ2)	no	yes	0.566	0.680	0.618	0.709
D (combined)	yes	yes	0.726	0.711	0.718	0.728

E. RQ2: Tracking, Filter-Free vs. Kalman

The second experiment compares a filter-free greedy-IoU tracker with a constant-velocity Kalman tracker using the same detector outputs.

Because RQ2 concerns temporal quality rather than detector recall, we evaluate bounding-box jitter and identity stability rather than using recall as the primary metric. The Kalman filter substantially reduces frame-to-frame bounding-box jitter. Mean centroid acceleration decreases by 32% on the unstabilized stream and by 43% on the stabilized stream.

The qualitative tracking results also show improved identity stability. In the filter-free configuration, a missed detection terminates a track, causing a subsequent detection to receive a new identity. The Kalman tracker can instead coast through a short detection gap using its predicted state and preserve the existing identity.

The Kalman-only configuration does not substantially improve detection recall, increasing recall only from 0.561 to 0.566. This confirms that the filter's primary contribution is not object recovery at the detector level. Most missed detections correspond to small or distant obstacles that were never observed by the detector, so a temporal filter cannot recover them.

Extending the coast period beyond one frame provides only limited additional recovery while increasing the risk of producing false "ghost" tracks. We therefore retain a one-frame coast horizon. Overall, the Kalman filter improves temporal smoothness and identity continuity rather than detection recall.

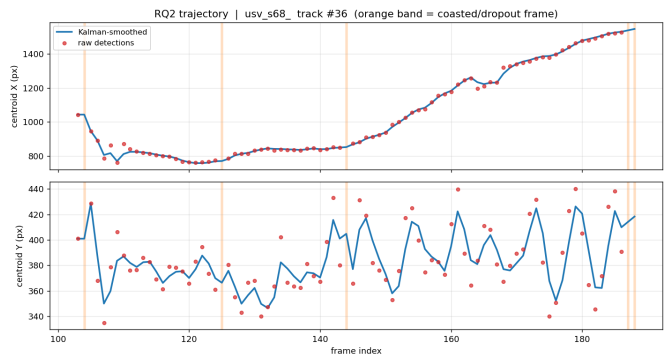

Fig. 2. Kalman smoothing of a track's centroid over time (USVTrack, one sequence). Raw per-frame detections (red) exhibit substantial frame-to-frame jitter; the Kalman-smoothed trajectory (blue) follows the true motion while suppressing it, corresponding to a 32–43% reduction in centroid acceleration. Shaded bands mark coasted (dropout) frames, where the track is propagated on its prediction alone.

Fig. 3. Filter-free vs. Kalman tracking on the same detections. Without a motion model (left) a missed detection ends a track and it re-appears with a new identity, yielding 71 identities over the sequence; the Kalman tracker (right) coasts through brief dropouts and preserves identity, yielding 60.

VI. DISCUSSION AND LIMITATIONS

A. Discussion.

The results show that horizon stabilization and temporal filtering address different failure modes in maritime perception.

For RQ1, camera motion is a substantial source of detection degradation. At the strongest evaluated perturbation, recall decreases from 0.891 on the clean input to 0.561 without stabilization. The synthetic IMU-guided stabilizer recovers recall to 0.723, recovering approximately half of the motion-induced loss. The amplitude sweep further shows that this benefit increases with motion severity, with only a small difference under mild motion and a substantially larger difference under severe motion.

For RQ2, the Kalman filter produces a different type of improvement. It does not significantly change detector recall, but it substantially reduces bounding-box jitter and improves identity continuity through short detection gaps. The improvement is particularly strong when the detector operates on stabilized frames, where the reduction in centroid acceleration increases from 32% to 43%.

The two modules are therefore complementary rather than simply additive. Stabilization addresses **detection degradation caused by camera motion**, while the Kalman filter addresses **temporal instability after detection**. Combining them produces the smoothest tracking output, but the modules do not compound substantially on recall because the Kalman filter cannot recover objects that were never detected.

B. Limitations.

Several limitations bound the conclusions of this study.

First, the stabilization pipeline is validated using synthetic IMU-derived motion rather than recorded IMU telemetry. The experiment therefore validates the geometric compensation mechanism under controlled motion and sensor-error assumptions but does not establish performance under real IMU drift, vibration, synchronization error, or camera-to-IMU frame misalignment.

Second, the camera-motion model is primarily geometric. It captures rotational camera motion but does not reproduce other effects associated with real vessel motion, including translational heave, motion blur, spray, glare, and other image-quality changes. Image warping can also introduce border artifacts.

Third, the stabilization and tracking experiments use the USVTrack validation split and five temporally correlated sequences. Consequently, these results should be interpreted as within-domain measurements rather than broad evidence of cross-domain generalization. The lower LaRS performance demonstrates that domain shift remains an important challenge.

Fourth, the trained detector focuses on dynamic obstacles because the shoreline-spanning static-obstacle category was excluded during dataset construction. Consequently, fixed hazards such as docks and other large static structures are outside the scope of the current detector.

Finally, jitter, dropout recovery, and identity counts are computed using the project's custom evaluation harness rather than standard multi-object tracking benchmarks. These metrics are useful for comparing the two tracking implementations internally, but they should not be interpreted as directly comparable to published MOT benchmark results.

VII. Summary and Future Work

This project investigated a class-agnostic maritime obstacle detection pipeline under camera motion and temporal uncertainty. A YOLOv8n detector was trained on a unified single-class corpus derived from USVTrack and LaRS, and two downstream mechanisms were evaluated: synthetic IMU-guided horizon stabilization and Kalman-based temporal tracking.

The stabilization experiment demonstrated that simulated camera pitch and roll can substantially degrade detection performance. At $\pm 30^\circ$ roll and $\pm 15^\circ$ pitch, recall decreased from 0.891 to 0.561, while the stabilization pipeline recovered recall to 0.723. The amplitude sweep further showed that stabilization becomes increasingly beneficial as camera motion becomes more severe.

The tracking experiment demonstrated that Kalman filtering addresses a different failure mode. Although it does not substantially improve detector recall, it reduces bounding-box jitter by 32% on the unstabilized stream and 43% on the stabilized stream while improving identity continuity through short detection gaps.

The next step is to replace the synthetic IMU-derived orientation estimates with synchronized real IMU telemetry. A dataset that pairs USV video with inertial measurements could provide a more realistic evaluation of the stabilization pipeline. Future work should also extend the motion model beyond pure rotational perturbations to include translational heave, motion blur, spray, and glare. Incorporating the currently excluded static shoreline and obstacle categories would broaden the detector's applicability to real navigation scenarios.

Acknowledgments

We thank Prof. Bruce Maxwell, Ha Yoon Kim, Claire Liu, and Tamir Yirga for their instruction and guidance throughout the course 'CS 5330: Pattern Recognition and Computer Vision', for which this project was undertaken.

References

[1] S. Han, D. Lee, J. Kim, and J. Kim, "A survey on maritime perception datasets and technologies for autonomous surface vessels," *Intelligent Service Robotics*, 2026.

[2] S. Yao, R. Guan, Y. Ni, S. Xu, Y. Yue, X. Zhu, and R. W. Liu, "USVTrack: USV-based 4D radar-camera tracking dataset for autonomous driving in inland waterways," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2025. arXiv:2506.18737.

[3] L. Žust, J. Perš, and M. Kristan, "LaRS: A diverse panoptic maritime obstacle detection dataset and benchmark," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2023, pp. 20304–20314.

[4] B. Bovcon, R. Mandeljc, J. Perš, and M. Kristan, "Stereo obstacle detection for unmanned surface vehicles by IMU-assisted semantic segmentation," *Robot. Auton. Syst.*, vol. 104, pp. 1–13, 2018.

[5] B. Bovcon and M. Kristan, "WaSR—A water segmentation and refinement maritime obstacle detection network," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 12661–12674, 2022.

[6] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[7] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2016, pp. 3464–3468.

[8] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960.