Occlusion-Aware UAV-Based Monitoring in Complex Environments: Applications to Infrastructure and Surveillance

Andreas Valvis
Department of Production and Management Engineering
Democritus University of Thrace (DUTH)
Xanthi, Greece
avalvis@pme.duth.gr

Vasiliki Balaska
School of Production and Management Engineering
Technical University of Crete (TUC)
and Institute of Computer Science, FORTH
Chania-Heraklion, Crete, Greece
vbalaska@ics.forth.gr

Loukas Bampis
Department of Electrical and Computer Engineering
Democritus University of Thrace (DUTH)
Xanthi, Greece
lbampis@ee.duth.gr

Antonios Gasteratos
Department of Production and Management Engineering
Democritus University of Thrace (DUTH)
Xanthi, Greece
agaster@pme.duth.gr

***Abstract*—Aerial inspection of complex infrastructure environments requires perception and control systems that can actively decide where to look next under uncertain and degraded visibility conditions. This paper presents an occlusion-aware UAV monitoring framework that integrates perception, navigation, and decision-making within a closed-loop inspection pipeline. The proposed approach couples prompt-guided pathway extraction, closed-loop route following, YOLOv11-based asset detection, overlap-driven occlusion reasoning, active viewpoint revisiting, and route-indexed counting within a unified inspection loop. The framework is primarily evaluated in the context of marine port inspection in an Unreal Engine/AirSim environment, while its underlying perception-to-indicator logic is further demonstrated in additional use cases, including border traffic monitoring and mining site inspection.**

***Index Terms*—UAV inspection, port monitoring, prompt guided segmentation, georegistration, object counting, critical infrastructure.**

I. Introduction

The autonomous monitoring of Unmanned Aerial Vehicles (UAV) is increasingly important for surveillance tasks in large areas of infrastructure environments [1], [2]. A practical system must operate with limited operator input, maintain a stable view of the inspection route, detect relevant objects, and transform observations into decision-support information. This work combines pathway-based monitoring with three dimensional localization, linking navigation, object detection, georegistration, and reporting within a unified framework.

The primary focus of this work is marine infrastructure inspection, which is used as the main application domain for the development and evaluation of the proposed approach. Ports constitute complex environments that combine service roads, quay edges, storage areas, vessels, buildings, cranes, restricted zones, and open water within a compact spatial layout. A key challenge is partial visibility caused by containers, cranes, parked vehicles, and berth structures, while maritime UAV imagery remains challenging due to clutter, scale variation, and scene complexity [3]. The proposed method addresses this by following visible pathway cues, detecting operationally relevant port assets with YOLOv11 [4], including vessels, vehicles, and personnel, estimating an occlusion risk score from image plane overlap, and revisiting ambiguous route segments from a modified viewpoint.

The main contributions of this work are as follows: (i) an occlusion-aware monitoring pipeline that integrates prompt-guided pathway extraction (CLIPSeg) [5], geometric route following, YOLOv11-based asset detection, and revisit-based counting within a closed-loop inspection process; (ii) a visibility-aware decision layer that uses image-plane overlap as an occlusion proxy and enables the active reacquisition of ambiguous targets from alternative viewpoints; and (iii) the extension of the same detection, georegistration, and indicator generation logic to additional use cases, including border traffic monitoring and mining site inspection, demonstrating the generality of the proposed framework.

II. Related Work

UAVs are widely used for infrastructure inspection and photogrammetric monitoring [1], [6], while marine environments introduce additional challenges, including moving vessels, service traffic, complex quay geometries, and restricted operating zones [2]. In this context, aerial detection benchmarks such as DOTA, oriented object detectors such as AO2-DETR,

This work was supported by the European Health and Digital Executive Agency (HADEA) under the MASTERMINE project (Grant No. 101091895).

Fig. 1. AirSim-based marine port simulation environment illustrating the proposed UAV inspection framework.

maritime datasets such as SeaShips, and efficient YOLO-based models address the small-object detection and cluttered scene conditions typical of port environments [3], [4], [7], [8].

Beyond detection, port monitoring requires the extraction of roads, quay edges, and docking pathways from visually similar surfaces. Prompt-guided models such as CLIPSeg enable text-conditioned pathway extraction [5], while tracking and geometric projection methods support the transformation of detections into stable, spatially consistent operational indicators [9]–[14]. However, existing approaches typically address these components in isolation. The integration of pathway following, occlusion-aware reasoning with active viewpoint revisiting, and consistent object counting within a closed-loop UAV inspection system remains insufficiently explored.

III. Methodology

A. Marine Port Monitoring

The marine system is organized as a closed-loop UAV inspection pipeline, operating within a simulated port environment, as illustrated in Fig. 1. The UAV climbs to a fixed surveillance altitude, streams imagery from a downward-facing camera, and treats the visible service road, lane marking, or docking pathway as the primary route cue. A front-facing view can be used for operator awareness; however, the route-following controller is driven by the downward stream.

Perception begins with road and pathway segmentation. The fast path uses HSV constraints to isolate grey road surfaces and white lane markings, followed by morphological closing, opening, and connected component filtering. When appearance is ambiguous, CLIPSeg is queried with prompts such as "road", "lane marking", "dock pathway", or "quay service road". Let I_t denote the RGB frame at time t, p the text prompt, and $M_t \in [0,1]^{H\times W}$ the pathway mask. The segmentation stage is:

$$M_t = S_\theta(I_t, p), \tag{1}$$

where $S_\theta(\cdot)$ is the prompt guided segmentation model with parameters θ, and H, W are image height and width, respectively. After thresholding and connected component filtering, the dominant pathway component yields center (c_x, c_y) and orientation ψ_t. The route following errors are:

$$e_x = \frac{c_x - \frac{W}{2}}{\frac{W}{2}}, \quad e_y = \frac{\frac{H}{2} - c_y}{\frac{H}{2}}, \quad e_\psi = \psi_t - \psi^\star, \tag{2}$$

where e_x is normalized lateral error, e_y is normalized forward centering error, and e_ψ is the yaw misalignment relative to desired pathway heading $\psi^\star$.

These errors drive a finite state behavior: search, center over the pathway, align with the pathway axis, settle, and follow. Velocity commands are smoothed and issued in the body frame, making the method independent of an absolute world heading and suitable for inspection pathways with changing orientation. The controller is designed for operational robustness: small unstable masks are rejected, the pathway must remain stable before forward motion, and the UAV can return to search when the road cue is lost.

The surveillance layer attaches object detection and counting to the route following behavior. A YOLOv11 detector produces port asset boxes B_i^t and confidences s_i^t for vehicles, ships, personnel, containers, and quay side equipment in frame t. Detections are associated temporally and spatially to suppress duplicates, then aggregated per route segment. Occlusion is not defined by a high IoU alone; rather, high image plane overlap is used here as an occlusion *proxy*. For target i, the overlap score is:

$$\rho_i^t = \max_{j\neq i} \frac{|B_i^t \cap B_j^t|}{|B_i^t \cup B_j^t|}, \tag{3}$$

where B_j^t are neighboring detections or projected obstacle regions, $|\cdot|$ denotes area, and $\rho_i^t \in [0,1]$. If $\rho_i^t > \tau_{\text{occ}}$ and $s_i^t < \tau_{\text{conf}}$, the target is marked visibility critical and the UAV revisits the corresponding route segment with a changed yaw or lateral offset to reduce overlap. Occlusion is not explicitly modeled; instead, image-plane overlap is used as a practical proxy for visibility degradation. The final output is a route-indexed inventory of uniquely counted assets, associated with port geography rather than individual frame observations.

B. Border Traffic Monitoring Use Case

To assess the transferability of the proposed framework, an additional use case is considered for border traffic monitoring. The UAV observes a road with a zoom-enabled RGB camera payload, while GNSS/IMU, gimbal state, camera calibration, and terrain information are used to convert detections into global coordinates (Fig. 2). The perception layer combines road segmentation with truck and vehicle detection. The road mask constrains the search region, reducing false detections outside the pathway, while the detector supplies truck proposals for queue analysis.

Queue tail estimation is handled through a pathway scanning policy. The road is discretized into segments, the gimbal

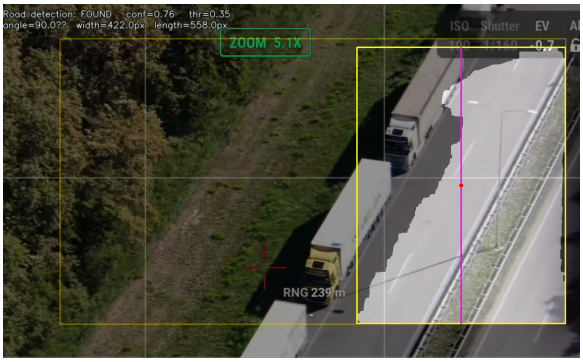

Fig. 2. Road pathway extraction and UAV-based traffic monitoring. The segmented road region defines the pathway of interest, while the estimated centerline provides a geometric reference for alignment. Vehicle detections within the pathway enable structured monitoring of truck traffic and support queue analysis under constrained observation.

Fig. 4. Mining site monitoring and 3D localization. (a) UAV-based object detection. (b) Projection of detections into a 3D point cloud. (c) Integration of localized objects into a georeferenced spatial representation.

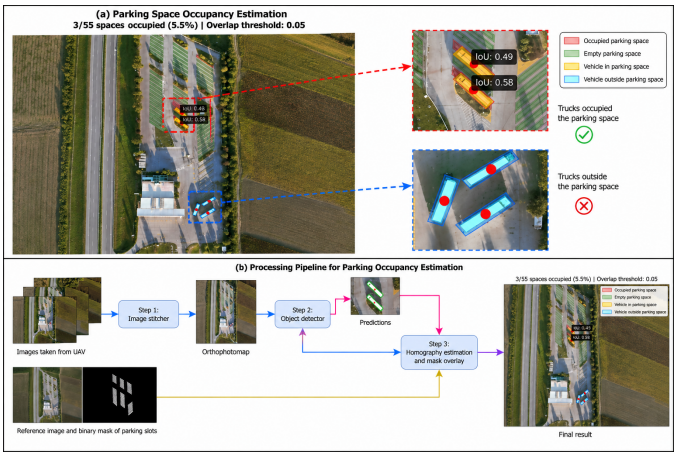

Fig. 3. (a) Parking occupancy estimation using UAV imagery, where vehicle detections are matched with predefined parking slots using an IoU-based criterion. (b) Processing pipeline including image stitching, object detection, and homography-based alignment with reference parking slot maps.

observes each segment centerline, and optical zoom is increased when necessary. Segments with a valid road mask but no truck detections are labeled truck free. The queue tail is defined as the most downstream detected truck before one or more truck-free segments. For georegistration, the center pixel of a truck detection is converted to a camera ray, intersected with a Digital Elevation Model (DEM) through a bounded Newton refinement, and optionally fused with a laser range finder return when available. The output is a geodetic queue tail coordinate with provenance metadata including the DEM source, geoid model, calibration profile, and residual. The parking module uses a complementary method. Drone frames are stitched when multiple nadir images are available; otherwise a single frame is processed directly. Vehicle detection uses oriented bounding boxes and sliced inference to preserve recall for small vehicles. The parking area is aligned to a reference orthophoto through homography estimation, and predefined parking slot polygons are projected into the drone image. Occupancy is decided by the Intersection Over Union (IoU) between each detected vehicle polygon and each slot polygon. The module returns total slots, occupied slots, vacant slots, occupancy rate, and per slot labels (Fig. 3). This is structurally close to port yard monitoring, where storage areas and restricted zones also require object-to-region assignment.

C. Mining Site Monitoring Use Case

Another use case is introduced to further evaluate the adaptability of the framework in mining environments. It detects heavy machinery and personnel, then localizes the detections in a 3D site representation. The implemented detector supports classes including bulldozer, car, driller, dump truck, excavator, grader, human, and truck. For each detection, the system utilizes the bounding box center, camera intrinsics, UAV pose, gimbal orientation, and a site point cloud. The point cloud is projected into the camera image, nearby projected 3D points are selected around the detection center, and the median 3D position is used as a robust object estimate.

Two localization modes are supported. In GNSS mode, metadata are read from DJI imagery or live telemetry, including latitude, longitude, absolute altitude, relative altitude, camera model, gimbal yaw, pitch, roll, and camera intrinsics. In visual localization mode, COLMAP based [12] pose estimation can replace direct GNSS metadata. A ground station bridge receives frames and telemetry through HTTP, stores synchronized JPEG and JSON records, runs object detection and 3D localization, and updates an Open3D point cloud viewer. The output is a structured record of detected object classes, image space boxes, UAV positions, object coordinates, and timestamps. The overall detection and localization pipeline is illustrated in Fig. 4, where image-space detections are projected on a 3D point cloud and integrated into a georeferenced site representation.

IV. Results

A. Marine Port Inspection

The marine port case is the main experimental contribution because it evaluates the full semantics to action loop rather

TABLE I
MARINE PORT RESULTS.

Metric	Score	Interpretation
Route completion	93.7%	End to end mission success
Mean lateral error	0.41 m	Pathway centering quality
Mean yaw error	5.6°	Route alignment stability
Count precision	0.90	False count control
Count recall	0.87	Asset recovery under clutter
Count F1 score	0.88	Unique inventory quality
Duplicate count rate	6.4%	Temporal association quality
Occlusion revisit success	81.3%	Visibility recovery
Runtime	17.8 FPS	Online feasibility

than an isolated detector. The UAV must first acquire the pathway, stabilize its heading, maintain the route under changing geometry, detect port assets, and then decide whether the current view is sufficient for counting or whether a revisit is required. Table I reports representative performance scores for the simulated port campaign.

Table I shows that the navigation and counting layers remain balanced rather than artificially perfect. Route completion at 93.7% indicates that the pathway follower is reliable across repeated inspection runs but still sensitive to difficult port geometry such as such as turns near docking/loading areas, stacked container shadows, and road-like concrete surfaces. The mean lateral error of 0.41 m and yaw error of 5.6° show that the prompt guided route representation is sufficiently stable for pathway surveillance, which is important since the subsequent counting stage depends on consistent camera placement rather than one off detections.

While the absolute count F1 score provides a useful performance indicator, the results highlight the significance of visibility-aware reasoning in improving detection consistency under occlusion. In cluttered marine scenes, assets are frequently partially hidden by cranes, containers, parked vehicles, and berth structures. A passive detector would simply miss those cases or count them inconsistently. Here, high overlap combined with low confidence marks a visibility critical event and triggers a revisit from a changed yaw or lateral offset. The 81.3% occlusion revisit success rate indicates that most ambiguous observations can be converted into usable counts after active reacquisition, while the duplicate count rate remains at 6.4%. These results indicate that the system can adapt its viewpoint to improve observation quality under challenging visibility conditions.

B. Border Traffic Monitoring Use Case

The border traffic results emphasize georegistration and decision support indicators. Table II summarizes the main results. Multiple highway recordings performed in Greece (Greek highways) and used as a preliminary field validation under controlled road conditions before testing in real border operation, while experiments performed in Serbia targeted operational queue tail and parking area behavior.

These results show that the same perception to indicator logic used in the marine case can also support transport monitoring when the required output is a geodetic queue tail coordinate or a parking occupancy estimate. The Greek highways provide a controlled field validation with mean localization accuracy of 98.29%, and the Serbian tests then extend the evaluation to queue tail localization and parking lot vehicle presence under operational conditions. Together, the results show that the framework can produce interpretable traffic indicators without changing the underlying architecture.

TABLE II
BORDER TRAFFIC AND PARKING RESULTS.

Scenario	Samples	Accuracy	Notes
Greek highways	21	98.29% mean	97.12–99.00% range
Serbian border	4	99.24%	Queue tail localization
Serbian parking lot	5	95.0%	Vehicle presence accuracy

C. Mining Site Monitoring Use Case

The mining site validation was performed in a relative field test designed to verify whether newly observed vehicles can be inserted into an already available georeferenced point cloud. Two vehicles were placed in the scene, while no corresponding vehicles were present in the reference 3D point cloud. In the tested scenario, the system successfully detected both vehicles, with an average confidence level of 87%, indicating that newly observed objects can be identified and associated with the existing spatial model without requiring full point cloud reconstruction for each mission.

This result is operationally important because it reduces both computational burden and delay in the monitoring workflow. Instead of regenerating a full georeferenced model every time a new flight is executed, the system can reuse the existing point cloud and localize newly detected vehicles directly within it. This makes the pipeline more suitable for repeated surveys and for near real time industrial monitoring, where fast integration of new observations is more valuable than repeated full reconstruction.

V. CONCLUSION

This paper formulates UAV-based inspection as an active perception problem, in which the system must determine when the current viewpoint is insufficient and autonomously acquire a more informative one. The proposed framework is developed and evaluated primarily within the context of marine port inspection, integrating prompt-guided pathway extraction, stable route following, YOLOv11-based asset detection, overlap-based revisit decisions, and route-indexed counting within a closed-loop inspection pipeline. Through this integration, aerial observations are transformed into structured, spatially consistent information at the port level. In addition, the application of the same perception-to-indicator logic to border traffic monitoring and mining site inspection provides evidence of the adaptability of the proposed framework across different operational environments, supporting its potential use in a broader range of infrastructure monitoring scenarios.

REFERENCES

[1] Y. Ham, K. K. Han, J. J. Lin, and M. Golparvar Fard, "Visual monitoring of civil infrastructure systems via camera equipped unmanned aerial vehicles: A review of related works," *Visualization in Engineering*, vol. 4, no. 1, 2016.

[2] X. Liu, J. Yang, and M. Zhou, "Deep learning based object detection in maritime unmanned aerial vehicle imagery: Review and experimental comparisons," *Applied Soft Computing*, vol. 151, 2024.

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

[4] G. Jocher and J. Qiu, "Ultralytics YOLO11," version 11.0.0, 2024. [Online]. Available: https://github.com/ultralytics/ultralytics

[5] T. Lueddecke and A. Ecker, "Image segmentation using text and image prompts," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 7086–7096.

[6] V. Balaska, I. T. Papapetros, K. M. Oikonomou, L. Bampis, and A. Gasteratos, "UAV Object Detection and Positioning in a Mining Industrial Metaverse With Custom Geo-Referenced Data," *IET Cyber-Physical Systems: Theory & Applications*, vol. 11, no. 1, Art. no. e70048, 2026.

[7] G. S. Xia *et al.*, "DOTA: A large scale dataset for object detection in aerial images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3974–3983.

[8] L. Dai, H. Liu, H. Tang, Z. Wu, and P. Song, "AO2 DETR: Arbitrary oriented object detection transformer," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 5, pp. 2342–2356, 2023.

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

[10] Y. Zhang *et al.*, "ByteTrack: Multi object tracking by associating every detection box," in *Proc. Eur. Conf. Comput. Vis.*, 2022.

[11] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[12] J. L. Schonberger and J. M. Frahm, "Structure from motion revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4104–4113.

[13] A. Telikani *et al.*, "Unmanned aerial vehicle aided intelligent transportation systems: Vision, challenges, and opportunities," *IEEE Communications Surveys & Tutorials*, 2025.

[14] Q. Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," arXiv:1801.09847, 2018.