

MULTI-CAMERA PANORAMIC SENSING FOR LOCALISATION AND MULTI-OBJECT TRACKING IN UAVS

Rafał Perz[1, 4], Paweł Woźniak[2], Rafał Chałupczyński[2], Konrad Barys[1, 4], Julian Konowalski[4], Artur Abratański[1, 4], Jakub Świerk[2] & Igor Dąbrowski[4]

[1]Warsaw University of Technology, Institute of Aeronautics and Applied Mechanics
[2]Uavionics Ltd Poland
[3]Warsaw University of Technology, Institute of Radioelectronics and Multimedia Technology
[4]Rafal Perz Scientific Network

Abstract

Perception-driven localisation and control constitute a core enabling capability for next-generation autonomous unmanned aerial vehicle (UAV) systems. This paper presents the project, with a hardware-synchronised omnidirectional perception system for UAVs, developed toward a 12-camera configuration and validated at this stage using a reduced multi-camera prototype. The system integrates panoramic reconstruction, multi-object detection and tracking, and vision-based geolocalisation for navigation in GNSS-denied environments. The paper reports initial and subsequent development results covering panoramic sensing, geometric–photometric calibration, detector – tracker optimisation, and optical payload integration

Keywords: Unmanned aerial vehicles, omnidirectional perception, multi-camera vision, vision-based localization, GNSS-denied navigation

1. Introduction

Perception-driven localization and control constitute a core enabling capability for next-generation autonomous unmanned aerial vehicle (UAV) systems operating in complex and congested airspace. The increasing density of UAV operations, combined with the limited reliability of global navigation satellite systems (GNSS) in urban, indoor, or contested environments, necessitates perception-centric approaches to navigation, traffic awareness, and separation assurance ([1], [2]).

Vision-based perception has emerged as a primary sensing modality for autonomous UAVs due to its extensive environmental information content and passive operation. Prior work has demonstrated the effectiveness of monocular and stereo vision systems for visual odometry, simultaneous localization and mapping (SLAM), and obstacle avoidance ([3], [4], [5]). However, most existing UAV perception pipelines rely on forward-facing or narrow field-of-view sensors, which inherently limit situational awareness and degrade performance in multi-agent or high-traffic scenarios [5]. Such configurations are particularly vulnerable to occlusions, rapid ego-motion, and delayed detection of lateral or rearward threats.

Omnidirectional and multi-camera sensing architectures offer a promising alternative by providing continuous 360° visual coverage. Previous studies have explored panoramic vision systems for ground robotics and autonomous driving, demonstrating improved robustness in localization and object tracking ([6], [7]). In the aerial domain, however, the integration of dense multi-camera arrays remains comparatively underexplored, primarily due to challenges related to synchronization, calibration,

computational complexity, photometric consistency between cameras, stitching artifacts, and real-time processing constraints on airborne platforms.

In parallel, recent advances in deep learning–based object detection and multi-object tracking have significantly improved the reliability of vision-based traffic perception. Convolutional detectors combined with tracking-by-detection paradigms have been successfully applied to aerial imagery and dynamic scenes ([8], [9]). Nevertheless, maintaining track continuity across wide fields of view and under aggressive platform motion remains an open problem, particularly in panoramic representations assembled from multiple sensors. This is especially relevant for UAV-based 360° perception, where object scale, inter-camera boundaries, exposure differences and local distortions may affect both detection confidence and identity preservation.

Another critical challenge concerns localization in GNSS-degraded or GNSS-denied environments. Vision-based localization methods leveraging epipolar geometry, feature-based mapping, and horizon-based cues have shown potential as standalone or complementary navigation solutions ([10], [11], [12]). The fusion of these methods within a unified omnidirectional perception framework offers an opportunity to improve robustness and operational continuity without reliance on external infrastructure.

Motivated by these challenges, this paper introduces project, a modular hardware–software omnidirectional perception system for UAVs based on a ring of hardware-synchronized cameras. The system integrates panoramic reconstruction, deep-learning-based object detection and tracking, and vision-based geolocalisation to support navigation and collision-risk assessment in GNSS-denied environments. The paper extends initial study by incorporating results obtained during the subsequent development stages focused on object detection and tracking algorithms, multi-camera image fusion, geometric–photometric calibration, and functional verification of the optical payload architecture.

2. Methods

2.1 Integrated panoramic reconstruction

For each calibrated camera, image pixels are back-projected into unit bearing vectors using the inverse intrinsic matrix. These vectors are transformed into a common UAV body (or global) reference frame using the camera rotation matrices. The resulting directions are mapped onto the unit sphere and expressed in angular coordinates, providing a unified spherical representation for panoramic image construction and multi-view alignment. This initial spherical reconstruction approach was subsequently refined during the next stage of the project by introducing a more detailed geometric and photometric calibration pipeline for the multi-camera payload. After spherical projection and geometric alignment, panoramic stitching is performed using optimisation based on a graph-cut formulation. Overlapping regions between adjacent views are represented as a graph in which pixels correspond to nodes and edges encode transition costs. An energy function combining data terms (view quality and viewing angle) and smoothness terms (intensity and gradient consistency) is minimised using a minimum cut / maximum flow algorithm. This yields an optimal seam that avoids high-gradient structures and reduces visible artifacts. The final panorama is generated by fusing images along the computed seam and applying multi-band blending to suppress exposure differences while preserving fine structural details.

In the implementation stage, the camera rig was additionally described by a combined geometric–photometric model. Each camera was characterised by its intrinsic parameters, lens distortion and extrinsic pose within the common multi-camera system. The calibration procedure also included focus adjustmen and colour correction, reducing sharpness, brightness and colour differences between neighbouring camera views. The refined reconstruction pipeline formulated image fusion as inverse reprojection onto a single virtual output camera, in which the viewing ray of each output pixel is projected into the calibrated image planes of the cameras observing that direction. This development did not replace the initial spherical stitching concept, but strengthened its calibration and preprocessing basis for subsequent panoramic detection, tracking and localisation experiments.

2.2 Detecting objects in real time

Multi-object tracking is performed using the BoT-SORT (Boosted Online Tracking and Re-identification) framework operating in a tracking-by-detection paradigm with YOLOv11 detections as input. In the subsequent development stage, the detection module was extended by evaluating additional YOLO-family configurations, including YOLOv26 variants, while the association stage was compared against ByteTrack in order to assess the trade-off between tracking accuracy, identity preservation and computational cost. BoTSORT extends the classical SORT approach by integrating appearance-based re-identification with motion-aware data association, improving identity preservation under occlusions, scale variations, and platform-induced motion. Each object track is represented by a Kalman filter state vector and an appearance embedding extracted from detected object regions. Motion prediction is used for spatial gating, while appearance similarity resolves ambiguous associations. ByteTrack was used as a lightweight association baseline relying primarily on geometric matching and detection confidence, whereas BoT-SORT was retained as the more robust target tracker for scenarios requiring stronger identity continuity.

Detection-to-track assignment is formulated as a bipartite matching problem with the following association cost:

$$C_{ij} = \alpha\left(1 - \mathrm{IoU}\left(b_i, \hat{b}_j\right)\right) + (1-\alpha)\ \| f_i - f_j \|_2 \quad (2)$$

where b_i denotes the detected bounding box, $\hat{b}_j$ is the predicted bounding box of track j, f_i and f_j are appearance embeddings, and α balances geometric and appearance cues. Optimal assignments are obtained using the Hungarian algorithm. Track lifecycle management includes confirmation and termination rules based on detection confidence and temporal persistence, reducing identity switches and false track fragmentation. During the later experimental stage, tracker parameters and detection thresholds were optimised on benchmark sequences before integration with panoramic UAV imagery, allowing the association module to be evaluated independently from the remaining panoramic reconstruction pipeline.

Tracking performance is evaluated using standard multi-object tracking metrics. Multiple Object Tracking Accuracy (MOTA) is defined as:

$$\mathrm{MOTA} = 1 - \frac{\sum_t (FN_t + FP_t + IDSW_t)}{\sum_t G\,T_t} \quad (3)$$

where FN, FP, and IDSW denote false negatives, false positives, and identity switches, respectively, and GT is the number of ground-truth objects/

Identity preservation is quantified using the Identity F1 score (IDF1), defined as

$$\mathrm{IDF1} = \frac{2 \cdot IDTP}{2 \cdot IDTP + IDFP + IDFN} \quad (4)$$

where IDTP, IDFP, and IDFN represent identity true positives, false positives, and false negatives. These metrics jointly assess detection reliability and long-term identity consistency across the panoramic field of view. These metrics are used to report the progress achieved after the subsequent detection and tracking experiments, rather than to indicate final validation of the complete perception system.

2.3 Navigation in GNSS-degraded environment

In GNSS-denied conditions, UAV localisation is formulated as a state propagation and correction problem initialised from the last valid GNSS measurement. The system maintains a state vector composed of position, velocity, and orientation, which is propagated using inertial measurements from onboard gyroscopes and accelerometers. Orientation is estimated through gyroscope integration, while translational motion is predicted via inertial dead reckoning. To constrain drift, the predicted pose is periodically corrected through visual alignment with a preloaded, georeferenced aerial map derived from satellite imagery. The last available GNSS fix provides a strong prior on the initial state, supporting visual–inertial correction during GNSS outages.

Visual map alignment provides a pose correction by minimising feature reprojection error:

$$\hat{\mathbf{p}}_t = \underset{\mathbf{p}}{\operatorname{argmin}} \sum_i \|\pi(\mathbf{T}(\mathbf{p}, \mathbf{q}_t)\mathbf{X}_i) - \mathbf{u}_i\|^2 \quad (5)$$

where $\mathbf{X}_i$ are georeferenced map features, $\mathbf{u}_i$ are observed image features, $\pi(\ )$ denotes the camera projection function, and $\mathbf{T}(\ )$ represents the pose transformation.

The visual correction is fused with the inertial prediction using a recursive filtering framework, enabling drift-limited localisation during extended GNSS outages.

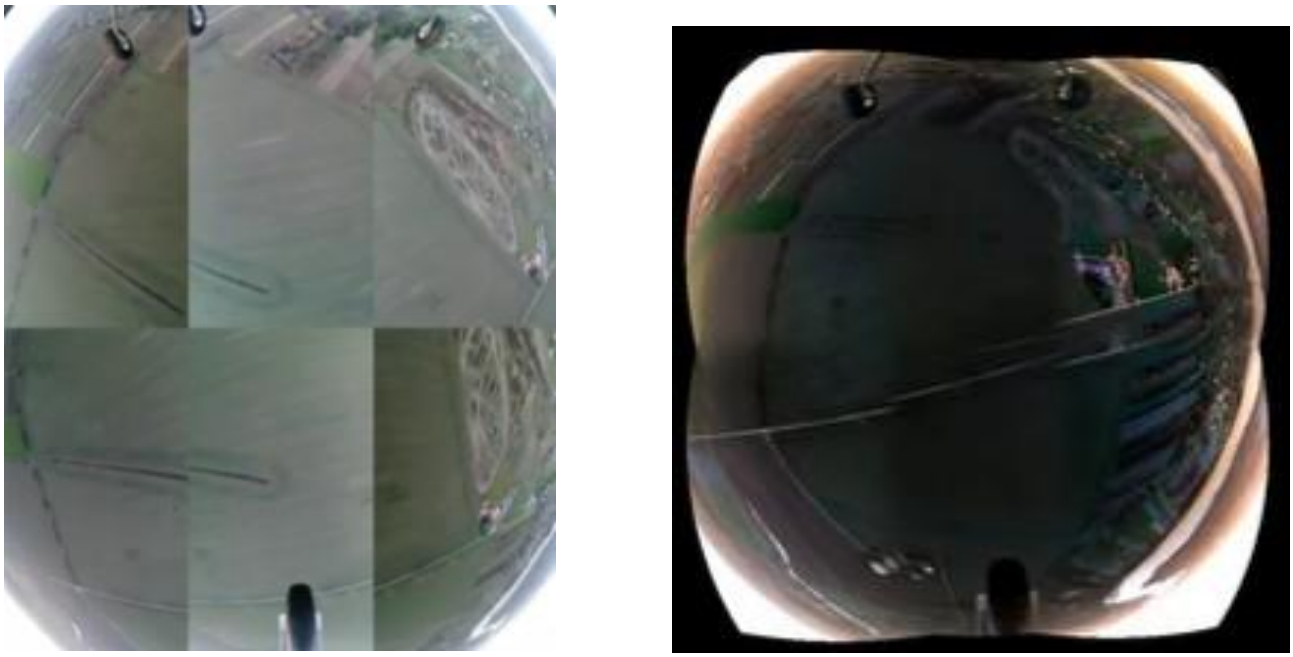
At this stage of the project, the localisation component is treated as a preparatory module integrated with the panoramic perception architecture rather than as a fully validated navigation subsystem. The current architecture of the integrated optical head supports the aggregation of video streams and detected objects from camera-dedicated processing units, together with carrier localisation data and onboard telemetry. Depending on the final integration configuration, telemetry data may be collected from the autopilot or from a dedicated IMU/GPS/MAG unit rigidly mounted with the cameras.

3. Results

3.1 Initial results

A functional prototype of the proposed omnidirectional perception system was implemented using a reduced hardware configuration comprising six hardware-synchronised cameras. An initial panoramic reconstruction pipeline was realised, in which individual camera views were geometrically aligned and projected onto a spherical surface. This spherical mapping enabled continuous omnidirectional scene representation and verified the feasibility of multi-camera stitching as a basis for downstream detection, tracking, and localisation tasks.

The detection and tracking module was evaluated using benchmarks from the MOTChallenge dataset. Performance was measured using Higher Order Tracking Accuracy (HOTA), Multiple Object Tracking Accuracy (MOTA), and Identity F1-score (IDF1). Three experimental configurations were considered: Dataset I (baseline tracking pipeline), Dataset II (enhanced data association and appearance modelling), and Dataset III (constrained configuration assessing robustness under reduced visual quality). All experiments were executed in a cloud-based environment using the RunPod platform equipped with an NVIDIA RTX 5090 GPU, ensuring consistent and reproducible evaluation conditions.



(a) The view from 6 cameras before stitching and overlay (b) Prototype view for opertator

Figure 1 – The results obtained from constructed protype with 6 cameras Basler 109204

Table 1 – Multi-object tracking performance on MOTChallenge sequences for three experimental configurations.

Sequence	Dataset I (Baseline)			Dataset II (Enhanced)			Dataset III (Constrained)		
	HOTA	MOTA	IDF1	HOTA	MOTA	IDF1	HOTA	MOTA	IDF1
Combined	**0.381**	**0.407**	**0.478**	**0.412**	**0.468**	**0.539**	**0.359**	**0.406**	**0.437**

The combined results indicate that the enhanced configuration (Dataset II), evaluated on the combined MOT15 and MOT17 benchmarks, yields consistent improvements across all metrics, with particularly pronounced gains in identity preservation (IDF1), reflecting improved robustness to occlusions and identity switches in denser and more heterogeneous scenes. In contrast, the constrained configuration (Dataset III), which further incorporates long-term tracking sequences from the PersonPath22 dataset, exhibits graceful performance degradation, indicating that while extended track continuity and long re-identification gaps increase task difficulty, the proposed system maintains stable tracking behaviour without catastrophic failure. In parallel, a dedicated dataset for GNSS-denied navigation was constructed, comprising georeferenced map data, synchronised multi-camera imagery, inertial measurements, and the last available GNSS fix. This dataset provides the basis for future evaluation of vision- and inertial-based localisation under GNSS outages.

3.2 Subsequent development results

Following the initial prototype validation, subsequent project tasks focused on refining the multi-camera reconstruction pipeline, improving detector–tracker performance, verifying real-time operation under field conditions, and extending the optical payload architecture toward a denser camera configuration.

The multi-camera calibration and image-fusion pipeline was refined through geometric and photometric modelling of the camera rig. The calibration process included focus adjustment, intrinsic and extrinsic parameter estimation, vignetting compensation and colour correction. These sub-pixel errors indicate very good geometric consistency of the camera rig, which is particularly important for panoramic stitching, since even small calibration deviations may cause visible discontinuities at the boundaries between neighbouring camera views.

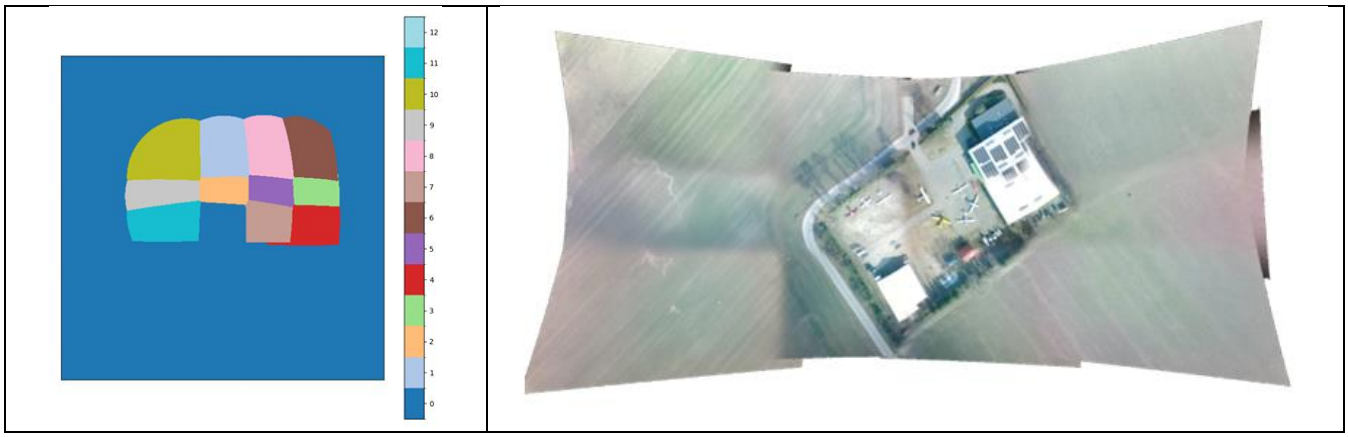


Figure 2 – Multi-camera reprojection results: (a) dominant-camera map; (b) final spherical panorama mapped onto a flat image plane.

The image-fusion pipeline was implemented as inverse reprojection onto a single virtual output camera. For each output pixel, the corresponding viewing ray was projected into the calibrated image planes of cameras observing that direction. Before reprojection, the input images were preprocessed through gamma removal, vignetting compensation and colour correction. Valid camera contributions were then combined using angle-based weighting, favouring observations closer to the optical axis, where distortion and calibration uncertainty are lower.

Detector and tracker development was continued through additional experiments with YOLO-family

configurations, including YOLO11 and YOLO26 variants. The experiments confirmed that enlarging and diversifying the training dataset had a stronger effect on tracking quality than increasing the number of epochs alone. The best YOLO11 configuration based on MOT15/17/20 achieved HOTA = 0.4124, MOTA = 0.4684 and IDF1 = 0.5390, while further YOLO26-based optimisation enabled higher tracking accuracy and identity preservation. The final aggregate result on the MOT15 benchmark reached MOTA = 0.501383 and IDF1 = 0.557889, slightly exceeding the milestone thresholds of MOTA > 0.50 and IDF1 > 0.55. Several individual MOT15 sequences achieved substantially higher values, confirming that the threshold was met at the aggregate level rather than through a single isolated result.

Model / configuration	Dataset	Epochs	HOTA	MOTA	IDF1
YOLO11n	MOT15/17	100	0.3806	0.4067	0.4783
YOLO11l	MOT15/17	100	—	0.4100	—
YOLO11l	MOT15/17/20	100	0.4124	0.4684	0.5390
YOLO11n	MOT15/17/20	200	0.3589	0.4063	0.4367
YOLO26n baseline	—	0	0.3547	0.3038	0.4221
YOLO26n	MOT15/17	50	0.4124	0.4684	0.5390

Table 2 – Selected detector–tracker results after transfer learning and model optimisation.

The computational-efficiency analysis showed a clear trade-off between detection quality and inference speed. Architectures with lighter backbones achieved higher throughput, whereas heavier or more expressive configurations provided better tracking metrics. The ResNet50 backbone combined with the YOLO26 head produced the strongest tracking results among the tested architectural modifications, while the MobileNet V3 backbone achieved the highest processing speed.

Model / configuration	Processing time [s]	FPS	MOTA	IDF1
YOLO26 skip connections	75.603	149.239	0.402035	0.493030
MobileNet V3 backbone + YOLO26 head	26.301	428.979	0.298535	0.388314
ResNet50 backbone + YOLO26 head	38.360	294.128	0.449284	0.512591

Table 3 – Selected computational-efficiency and tracking-quality comparison.

Field tests were conducted during UAV flights to verify real-time operation of the detection software under practical conditions. These tests were carried out using the earlier YOLOv11-based pipeline, prior to migration to the YOLO26 detector. The YOLOv11-based detection pipeline operated stably on the onboard Radxa ROCK 3E computer, and no significant delay was observed between image acquisition and generation of detection outputs. The tests indicated that the main limitation was not processing speed alone, but the mismatch between public training datasets and real UAV imagery, especially in terms of viewing perspective, object scale, flight altitude and illumination conditions.

A dedicated 360° image case study showed that panoramic detection quality is affected by local distortions, exposure differences, snow-dominated backgrounds and visible seams between camera sectors. The algorithm missed some buildings, vehicles and road segments visible in the panorama, particularly near visually difficult regions and sector boundaries. These results indicate that further improvement requires exposure normalisation between cameras, seam-aware fusion of detections and dedicated training data for panoramic UAV imagery.

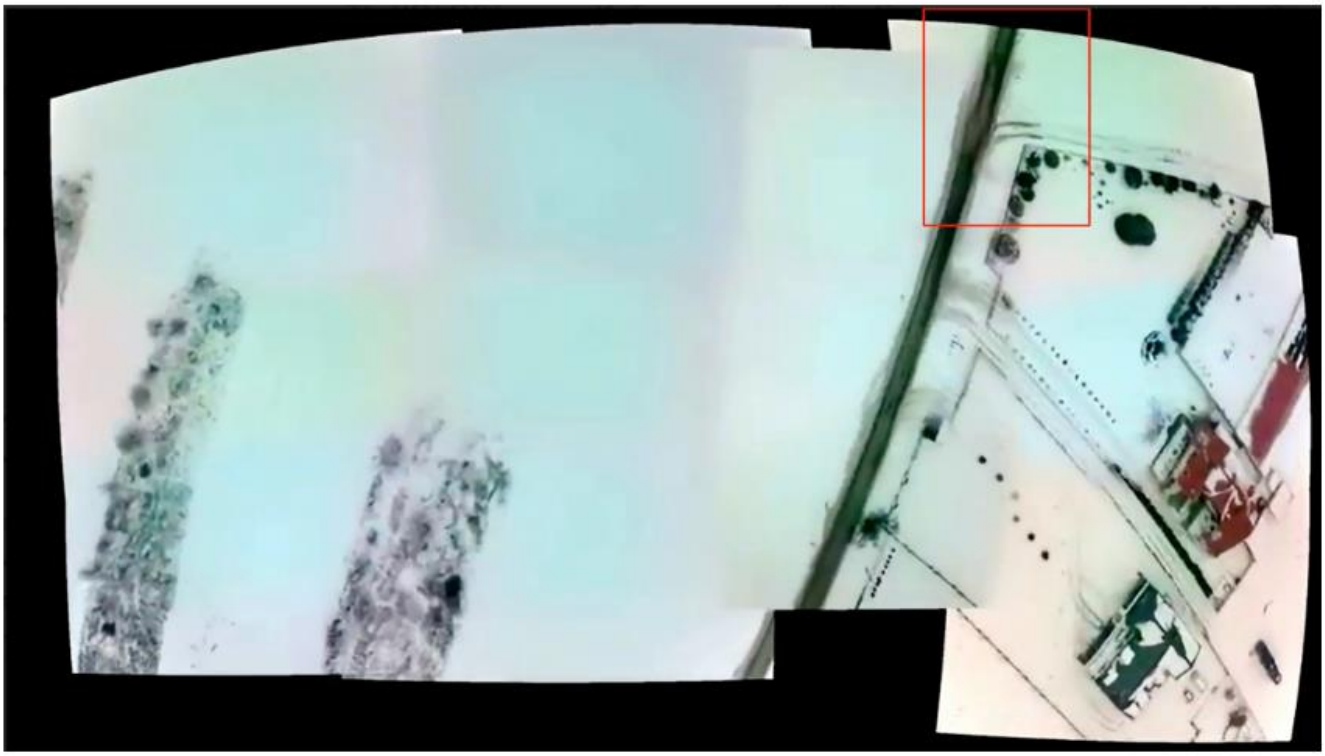


Figure 3 – Example of object detection on a 360° panoramic image.

In parallel with algorithmic development, the optical payload was extended toward a denser camera architecture. The functional integrated optical head concept assumes distributed camera-dedicated processing units connected to an aggregator through a local Ethernet network. The aggregator combines video streams, detection outputs, UAV telemetry and localisation data, and provides an operator interface. The optical head was further developed toward a 12-camera array supported by an additional wide-field reference camera, which separates high-resolution perception streams from a low-latency operator overview and provides the basis for subsequent end-to-end panoramic perception tests.

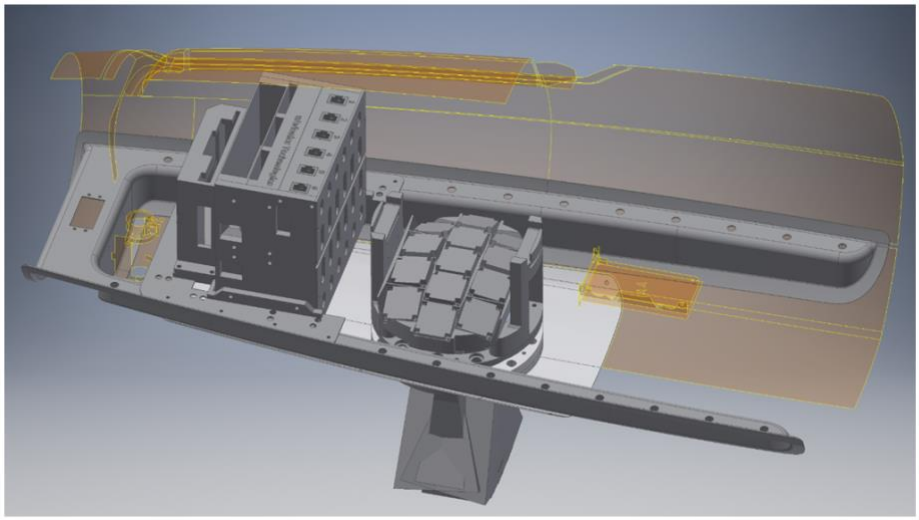


Figure 4 – Functional model of the current camera matrix installed in the UAV payload bay.

The subsequent development results confirm that the system evolved from an initial six-camera functional prototype toward a more mature multi-camera perception payload. At this stage, the work demonstrates improved calibration, image-fusion capability, detector–tracker optimisation and functional optical-head integration, while full end-to-end validation of localisation and autonomous

response functions remains a further project stage.

4. Conclusions

This work presented an initial prototype of an omnidirectional multi-camera perception system for uncrewed aerial vehicles, integrating calibrated panoramic reconstruction, hardware-supported multi-stream video processing, multi-object detection and tracking, and preparatory components for GNSS-denied localisation. A six-camera hardware prototype was implemented and used to validate the feasibility of multi-view synchronization, geometric and photometric camera calibration and spherical panoramic mapping. The detection and tracking pipeline was quantitatively evaluated on MOT15, MOT17, and selected extended tracking configurations, demonstrating stable performance and graceful degradation under in-creasing scene complexity. The optimised detector–tracker configuration confirmed the feasibility of reaching the assumed tracking quality thresholds, while additional 360° and field-oriented tests indicated that panoramic UAV imagery requires dedicated handling of scale variation, camera-sector boundaries, exposure differences, local distortions, and seam-aware fusion of detections.

The results also confirmed the relevance of a heterogeneous hardware–software architecture for the proposed payload. FPGA/MPSoC-based tests demonstrated that multi-stream acquisition, preprocessing, aggregation, and selected computer-vision functions can be supported by a dedicated low-latency processing layer. This reduces the burden on the main perception pipeline and supports further scaling toward denser camera configurations and real-time onboard operation.

In the next phase, dedicated experimental campaigns will be conducted to evaluate vision-based geolocalisation under GNSS-denied conditions using the constructed map- and inertial-based dataset. The perception system will be further validated in a twelve-camera configuration supported by an additional wide-field reference camera, separating high-resolution perception streams from a low-latency operator overview. This architecture will enable the separation of high-resolution perception streams from a low-latency operator overview and will support further integration of the panoramic reconstruction, detection, tracking and localisation modules. Detection capabilities will be extended to explicitly identify other UAVs in the airspace and to estimate relative motion parameters for the execution of avoidance manoeuvres. In parallel, ground-scene perception will be enhanced by integrating vehicle detection to support comprehensive situational awareness. These steps define a structured roadmap toward full-scale autonomous perception and navigation tests in operational UAV scenarios.

The subsequent development stage confirmed that the initial six-camera prototype provided a suitable basis for further development of the Sauron Eye omnidirectional perception payload. The later work extended the system beyond the initial feasibility demonstration by improving geometric and photometric calibration, refining the multi-camera image-fusion pipeline, optimising detector–tracker configurations, and developing the optical head toward a denser multi-camera architecture.

The calibration and image-fusion results demonstrated that the multi-camera rig can be modelled with sufficient geometric accuracy for panoramic reprojection. The achieved reprojection error confirmed the suitability of the calibrated camera model for further panoramic image generation, while vignetting compensation and colour correction reduced visual inconsistencies between neighbouring camera views. These results indicate that the reconstruction pipeline evolved from an initial spherical stitching concept toward a more robust geometric–photometric processing chain.

The detector and tracker experiments showed that tracking performance depends not only on the selected neural-network architecture, but also on the training dataset, association strategy and tracker hyperparameters. The best evaluated configurations improved MOTA and IDF1 values compared with the initial setup, confirming that further optimisation of the detection and association stages is necessary for reliable panoramic UAV perception. At the same time, field tests indicated that real UAV imagery introduces limitations not fully represented in public benchmark datasets, particularly in terms

of viewing angle, object scale, illumination and panoramic distortions.

The 360° image case study further showed that detection quality is affected by image-level properties of the panoramic representation, including local distortions, exposure differences and visible transitions between camera sectors. This indicates that future improvements should not be limited to detector training alone, but should also include seam-aware preprocessing, improved photometric normalisation and training data dedicated to panoramic UAV imagery.

The hardware-oriented work confirmed the feasibility of extending the optical payload toward a multi-camera integrated optical head architecture with distributed camera processing units and an aggregator responsible for combining video streams, detections, telemetry and localisation data. The development toward a 12-camera array supported by an additional wide-field reference camera provides a more mature hardware basis for future end-to-end panoramic perception tests.

Overall, the subsequent development results confirm measurable progress in calibration, image fusion, detection, tracking and payload integration, while the full validation of GNSS-denied localisation and autonomous response functions remains a further stage of the project. The next phase should therefore focus on integrated tests of panoramic reconstruction, multi-object tracking and vision–inertial localisation in realistic UAV flight scenarios, using dedicated datasets and the extended multi-camera payload configuration.

5. Acknowledgments

The project was a result of cooperation between UAvionics Ltd and Rafal Perz Research Network within the funds from the program Ścieżka SMART from EU funds.

Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

References

[1] Shlomo Tsach, Akiva Peled, David Penn, Ben Keshales, and R Guedj. Development trends for next generation of uav systems. In *AIAA Infotech@ Aerospace 2007 Conference and Exhibit*, page 2762, 2007.

[2] Hüseyin Küçükerdem, Cemal Yilmaz, Hamdi Tolga Kahraman, and Yusuf Sönmez. Autonomous control of unmanned aerial vehicles: applications, requirements, challenges. *Cluster Computing*, 28(11):734, 2025.

[3] Vicky S Kalogeiton, Konstantinos Ioannidis, G Ch Sirakoulis, and Elias B Kosmatopoulos. Real-time active slam and obstacle avoidance for an autonomous robot based on stereo vision. *Cybernetics and Systems*, 50(3):239–260, 2019.

[4] Abhishek Gupta and Xavier Fernando. Simultaneous localization and mapping (slam) and data fusion in unmanned aerial vehicles: Recent advances and challenges. *Drones*, 6(4):85, 2022.

[5] Asma Hamissi and Amine Dhraief. A survey on the unmanned aircraft system traffic management. *ACM Computing Surveys*, 56(3):1–37, 2023.

[6] Muhammad Yeasir Arafat, Muhammad Morshed Alam, and Sangman Moh. Vision-based navigation techniques for unmanned aerial vehicles: Review and challenges. *Drones*, 7(2):89, 2023.

[7] Dawei Du, Yuankai Qi, Hongyang Yu, Yifan Yang, Kaiwen Duan, Guorong Li, Weigang Zhang, Qing-ming Huang, and Qi Tian. The unmanned aerial vehicle benchmark: Object detection and tracking. In *Proceedings of the European conference on computer vision (ECCV)*, pages 370–386, 2018.

[8] Hongyang Yu, Guorong Li, Weigang Zhang, Qingming Huang, Dawei Du, Qi Tian, and Nicu Sebe. The unmanned aerial vehicle benchmark: Object detection, tracking and baseline. *International Journal of Computer Vision*, 128(5):1141–1159, 2020.

[9] Danilo Avola, Luigi Cinque, Anxhelo Diko, Alessio Fagioli, Gian Luca Foresti, Alessio Mecca, Daniele Pannone, and Claudio Piciarelli. Ms-faster r-cnn: Multi-stream backbone for improved faster r-cnn object detection and aerial tracking from uav images. *Remote Sensing*, 13(9):1670, 2021.

[10] Omar Y Al-Jarrah, Ahmed S Shatnawi, Mohammad M Shurman, Omar A Ramadan, and Sami Muhaidat. Exploring deep learning-based visual localization techniques for uavs in gps-denied environments. *IEEE Access*, 2024.

[11] Mitch Bryson and Salah Sukkarieh. A comparison of feature and pose-based mapping using vision, inertial and gps on a uav. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4256–4262. IEEE, 2011.

[12] Chun Liu, Shuhang Zhang, and Akram Akbar. Ground feature oriented path planning for unmanned aerial vehicle mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(4):1175–1187, 2019.