

EVPeriscope: Extended Perception across Aerial and Ground Vehicles with Event-based Propeller Tracking

Dexter Ong, Vijay Kumar, and Pratik Chaudhari

General Robotics, Automation, Sensing and Perception (GRASP) Laboratory,
University of Pennsylvania, Philadelphia PA 19104, USA

Abstract. Reliable relative localization between aerial and ground robots is a key requirement for tightly coordinated heterogeneous teams. This can be difficult to do using conventional frame-based cameras and fiducial markers because they are sensitive to motion blur, lighting variations, and payload constraints. This paper presents EVPeriscope, an event-based perception system that enables detection, localization and control of a quadrotor using an upward-facing event camera on a ground robot by detecting the high-frequency visual signature of its propellers. This system allows the quadrotor to function as an extended perception system for the ground robot when onboard sensors exhibit degradation or occlusion. We demonstrate the capabilities of this marsupial ground-aerial system via experiments in challenging field conditions with wind speeds of up to 15 mph, in both daylight and at night. We show that the system supports localization and closed-loop navigation through dense foliage where the ground robot’s sensors are occluded. Our control system for the quadrotor operates at 200 Hz entirely with onboard sensing and computation. More details and experiment videos can be found on the project page: <https://ongdexter.github.io/evperiscope>.

1 Introduction

Heterogeneous aerial-ground robot teams are increasingly deployed across a wide range of applications, including search and rescue [1], infrastructure inspection [2], and precision agriculture [3]. These domains require robots that can rapidly adapt their sensing perspective, cover large areas efficiently, and operate reliably under adverse conditions. Aerial-ground robot teams are particularly well-suited to meeting these demands, as they combine complementary sensing and mobility. An unmanned ground robot (UGV) can provide endurance, computation, and payload capacity, while an unmanned aerial vehicle (UAV) can rapidly obtain elevated viewpoints over obstacles, clutter, and confined terrain. Realizing such behaviors, however, requires reliable relative localization between the two robots under fast motion, changing illumination, and onboard computing constraints.

While position estimates provided by a global positioning system (GPS) module may be sufficient if the landing zone is large, vision-based systems can provide higher positioning accuracy for precision landing. One option is to mount fiducial markers (e.g., AprilTags [4]) on the UAV and track them with a conventional camera. For small quadrotors, this introduces a practical tradeoff: markers must be large enough to detect reliably, yet they add weight, occupy limited surface area, and may interfere with airflow around the vehicle [5]. A fiducial marker could be placed on the landing zone [6, 7] but

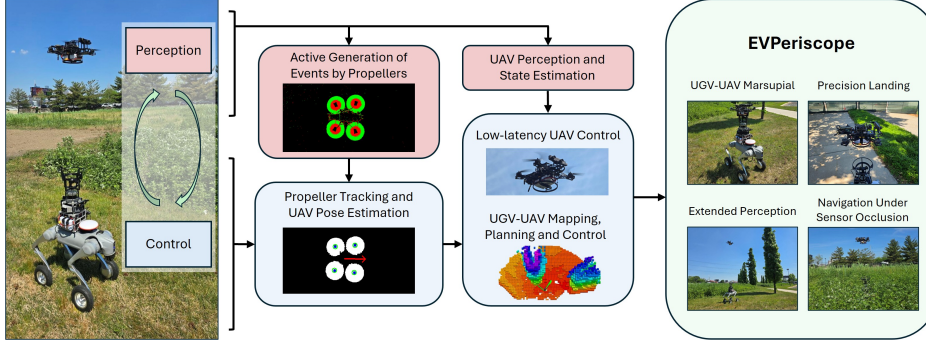


Fig. 1: EVPeriscope is a marsupial ground-aerial system. It combines the endurance and computational capability of the UGV and the elevated field-of-view of the UAV. It uses an upward-facing event camera on the ground robot to capture the high-frequency spatiotemporal structure of the propellers of the quadrotor. The propellers are detected, tracked and used for low-latency pose estimation and control of the UAV at 200 Hz. This enables precision landing of the UAV on the small footprint of the ground robot which is crucial for long-range deployment of the marsupial system. Tightly coupled localization and control between the aerial and ground robot allows the use of the UAV sensors as an extended perception system, enabling behaviors such as mapping and navigation in environments where the UGV sensors are deteriorated or occluded.

this requires the quadrotor to have an additional downward-facing camera just for takeoff and landing. In addition, the same size constraints on the size of the fiducial marker also apply to smaller landing zones such as small ground robots. These limitations motivate a markerless approach that uses visual structure already present on the aerial vehicle.

Event cameras are well-suited to this setting because they can capture rapid brightness changes at a microsecond temporal resolution and have a high dynamic range which enables robust operation under extreme lighting changes. In contrast, even high-speed frame-based cameras suffer from motion blur, limited dynamic range, and degraded performance in low-light or high-contrast conditions. Spinning propellers produce strong, localized event activity that remains detectable even when conventional imaging fails. We leverage this to estimate propeller locations and, using their known geometry, recover the quadrotor pose for relative localization and control between the ground and aerial robot. This enables extended perception: the UAV acts as a periscope to augment the UGV’s sensing.

An overview of the EVPeriscope system is provided in Fig. 1. This paper presents a lightweight event-based pipeline for quadrotor localization and control using an upward-facing event camera on a ground robot. This enables the use of the quadrotor as an extension of the ground robot’s perception system to augment its sensing capabilities. The contributions of this paper are as follows.

1. A real-time markerless event-based relative localization system enabling closed-loop UAV tracking from an upward-facing UGV-mounted event camera at 200 Hz under challenging outdoor conditions.
2. A ground-aerial marsupial system with a tightly coupled extended perception architecture in which the UAV functions as an extendable sensor for the UGV like a periscope.

3. Field validation demonstrating robust operation in dense foliage, nighttime conditions, and wind disturbances using fully onboard sensing and computation.

2 Related Work

2.1 Event-based UAV Detection

Event-based propeller detection leverages neuromorphic cameras to exploit the fast, periodic motion of UAV propellers, enabling robust detection and tracking of UAVs under challenging lighting and background clutter. EVPropNet [5] models propeller geometry to generate synthetic event streams; event data is converted into an image for propeller detection by a convolutional neural network and used for UAV detection, tracking and landing. Spiking neural networks have also been used for detection of UAVs in low-power applications [8]. More recent pipelines like EventPro [9, 10] focus on dynamics by estimating propeller rotational speed via motion-compensated event batches.

With the proliferation of accessible and low-cost consumer UAVs, UAV detection is a crucial capability for autonomy in robotics, but also for safety and security. Multiple datasets have been created to help drive research in event-based UAV detection [11–13]. Event-based detection of UAVs has also been proposed for the application of virtual fences in surveillance [14]. We use a simple and lightweight approach for detection and tracking of propellers from a ground robot to enable low-latency localization and control of the quadrotor. In contrast to learning-based detectors, this representation operates directly on the incoming events and requires neither a trained model nor propeller-specific training data. We present an application of event-based UAV detection for persistent relative localization of the full pose of the UAV for tightly coupled heterogeneous coordination.

2.2 UAV Landing Systems

GPS can provide position information to aid autonomous landing of UAVs [15]. However, without RTK correction which is often unavailable in the field, GPS can have significant noise, on the order of meters, which is inadequate for precision landing. Ultra-wideband positioning systems have been used to achieve positioning accuracy on the order of tens of centimeters [16]. Even so, when the landing zone is small due to space constraints, such as on a small ground robot, this accuracy is insufficient. Vision-based landing systems [17–19] localize the landing pad using fiducial markers which can provide more accurate 6-degree-of-freedom (6-DoF) pose estimates when the marker is clearly visible. However, these methods are constrained by the size and capacity of both the ground and aerial robots, requiring a marker or downward-facing camera mounted on the UAV. They also suffer from limitations of conventional frame-based sensors, resulting in degradation of detection performance under adverse illumination conditions and motion blur [20, 21]. We exploit neuromorphic perception to perform robust, target-less and low-latency quadrotor detection and tracking, leveraging the fact that quadrotor propellers actively generate a distinctive high-frequency temporal signature in the event stream.

2.3 Marsupial Ground-Aerial Systems

Marsupial robot teams use a carrier robot to transport, deploy, recover, or support one or more child robots, combining the endurance and payload capacity of a larger platform with the mobility and viewpoint diversity of smaller agents [22]. Most approaches utilize the UAV to achieve a complementary objective with the UGV for exploration and mapping [23–25]. These approaches typically deploy the UAV from the UGV once per mission and do not account for recovery and multiple deployments. Instead of treating the marsupial system as independent robots after the initial deployment, this work explores the use of the UAV as a tightly integrated extension of the UGV perception system where the UAV is dynamically deployed and recovered.

3 Method

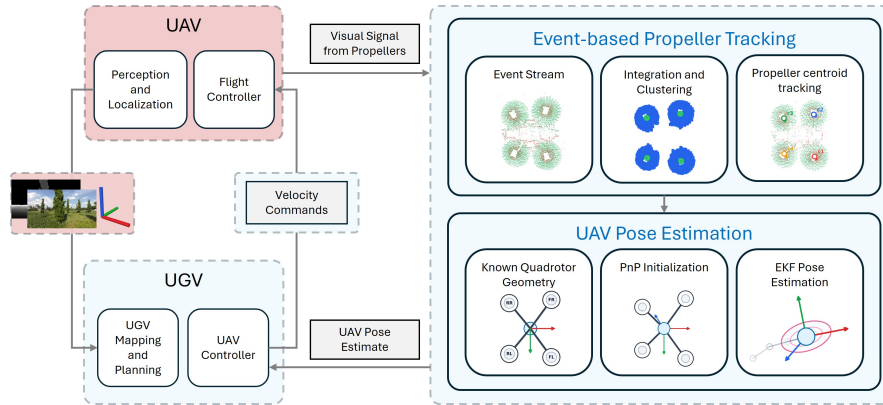


Fig. 2: EVPeriscope system diagram. The UGV-mounted event camera detects and tracks the UAV propellers from asynchronous events, clusters them into propeller centroids, and uses the known asymmetric quadrotor geometry to estimate the UAV pose. The estimated pose is then used by the UGV controller to command the UAV through velocity commands. UAV observations and state estimates are transmitted to the UGV as extended perception that can be used for UGV autonomy.

The EVPeriscope system consists of an event-based propeller detection and tracking pipeline, quadrotor pose estimation module and UAV controller. The system diagram is shown in Fig. 2.

3.1 Event-based Propeller Detection and Tracking

Event cameras produce sparse, high-temporal-resolution measurements that are inherently asynchronous and noisy, making instantaneous frame-based detection unreliable. To robustly capture the spatiotemporal structure of rapidly rotating propellers against noise and background events generated by the environment, we maintain a continuously updated state that emphasizes recent events while suppressing older ones.

The propeller detector maintains a persistent state “image” $S(u, v)$ with $u, v \in \mathbb{Z}$ that is updated once per event batch using a first-order exponential IIR filter on the event stream to result in leaky temporal integration. At the end of each batch, the previous state is decayed by a global factor. Each event contributes to the pixel where it occurred with a weight that depends on how recently it occurred. Altogether,

$$S(u, v) \leftarrow S(u, v) \exp\left(-\frac{(t_{\text{end}} - t_{\text{last}})}{\tau}\right) + \sum_{i=1}^n \exp\left(-\frac{(t_{\text{end}} - t_i)}{\tau}\right) \delta_{u, u_i} \delta_{v, v_i}. \quad (1)$$

where $S(u, v)$ denotes the filter state at pixel (u, v) , the exponential time constant is τ , the end time of the current batch is t_{end} , end time of the previous batch is t_{last} , the number of events in the batch is n , and t_i is the timestamp of event i . The Kronecker delta function is non-zero only when $u = u_i$ and $v = v_i$ for the event pixel (u_i, v_i) . To maximize response to the fast-moving propellers, all events are treated as positive contributions regardless of their polarity. We process the event stream in 5 ms windows, i.e., $t_{\text{end}} - t_{\text{last}} = 5$ ms, and decay the representation with $\tau = 20$ ms. These parameters were fixed across all experiments.

Pixels in S with intensity one standard deviation above the mean (both computed only on pixels with non-zero intensities) are used to obtain a binarized image. This image is cleaned using morphological erosion, followed by dilation, to suppress background noise. The k -means algorithm is used to obtain candidate centroids on this cleaned image. With $k = 4$, we obtain clusters corresponding to events from each of the propellers. We initialize the k -means algorithm using centroids detected in the previous image. This technique can, in principle, be extended to an N -rotor platform by setting $k = N$. The persistent state ensures that our detector for the centroids is stable over time and clusters depict clear structure in the propellers. The tracker maintains temporal continuity by initializing the estimates of the centroids from the previous frame and refining them using the regions that are detected in the current frame.

3.2 Quadrotor Pose Estimation

We assume that the quadrotor geometry is known and that the frame is asymmetric, with the front arms differing in length and relative angle from the rear arms. This asymmetry allows us to use the centroids of the propellers detected in the previous section to estimate the full 6-DoF pose of the quadrotor. If the quadrotor is symmetric or its geometry is unknown, the position of the quadrotor can still be estimated for tracking and landing, similar to what is done in prior work [5]. However, yaw would not be observable in such cases. Geometric correspondence is established in our case by ordering the centroids. Let the centroid of the i^{th} propeller in the image plane be $\mathbf{z}_i \in \mathbb{R}^2$ and denote

$$\mathbf{z} = [\mathbf{z}_1^\top, \mathbf{z}_2^\top, \mathbf{z}_3^\top, \mathbf{z}_4^\top]^\top.$$

Let the mean be $\bar{\mathbf{z}} = \frac{1}{4} \sum_i \mathbf{z}_i$ and their orientation with respect to the mean be

$$\gamma_i = \tan^{-1} \left(\frac{(\mathbf{z}_i - \bar{\mathbf{z}})_y}{(\mathbf{z}_i - \bar{\mathbf{z}})_x} \right).$$

For our platform, the front pair of propellers has the largest distance between them. In this work, we require visibility of all four propeller centroids, although in principle three propellers would be sufficient. Since we know the propeller locations in the quadrotor body frame exactly, an initial pose is obtained using Perspective-n-Point (PnP). This estimate is updated with a discrete-time Extended Kalman Filter (EKF). The state is

$$\mathbf{x} \equiv [\mathbf{p}^\top, \boldsymbol{\Theta}^\top, \mathbf{v}^\top, \boldsymbol{\omega}^\top]^\top \in \mathbb{R}^{12},$$

where $\mathbf{p} \in \mathbb{R}^3$ is the quadrotor position in the camera frame, $\boldsymbol{\Theta} = [\phi, \theta, \psi]^\top$ is the attitude, and $\mathbf{v}, \boldsymbol{\omega} \in \mathbb{R}^3$ are the translational and angular velocities. We assume constant velocity dynamics. Observations are given by the propeller centroids $\mathbf{z} \in \mathbb{R}^8$ in the image plane. For the i^{th} propeller, let

$$\mathbf{q}_i = [X_i \ Y_i \ Z_i]^\top \in \mathbb{R}^3$$

denote its position in the camera frame. The corresponding image measurement is

$$\mathbf{z}_i = \mathbf{h}_i(\mathbf{x}) \equiv \begin{bmatrix} c_x + f_x \frac{X_i}{Z_i} \\ c_y + f_y \frac{Y_i}{Z_i} \end{bmatrix}.$$

Here, $\mathbf{q}_i$ depends on the quadrotor attitude $\boldsymbol{\Theta}$, position $\mathbf{p}$, and known propeller geometry. The camera focal lengths are (f_x, f_y) in pixels, and (c_x, c_y) denotes the principal point.

3.3 UAV Control

Given the estimated UAV pose from the EKF, we generate UAV body-frame velocity commands as follows. The event camera is aligned with the body frame of the ground robot. In this work, we set the target state for the quadrotor to be centered on the camera with the yaw aligned with the front of the ground robot. We therefore design a UAV velocity controller to track the error in position $\mathbf{p}$ and yaw ψ in the camera frame of the ground robot. For each axis, a standard PID controller determines the control, with separate gain sets for planar motion, vertical motion, and yaw. This velocity is transformed into the frame using the estimated yaw in the EKF before publishing to the UAV. This ensures that the controller output is consistent with the vehicle command interface while still regulating position in the camera reference frame.

3.4 UAV Takeoff and Landing

The UAV is first commanded to take off to a set point of 1.5 m above the ground robot. EVPeriscope is then initialized on the UGV to start the propeller detection and tracking modules. Upon stable localization of the UAV, the EVPeriscope controller begins streaming velocity commands to the UAV. The quadrotor is then transitioned to a tracking state and control authority is handed off to the UGV.

When the landing sequence is triggered, the UAV is commanded to a height of 0.5 m. To ensure landing under stable wind conditions, the controller checks that the UAV is

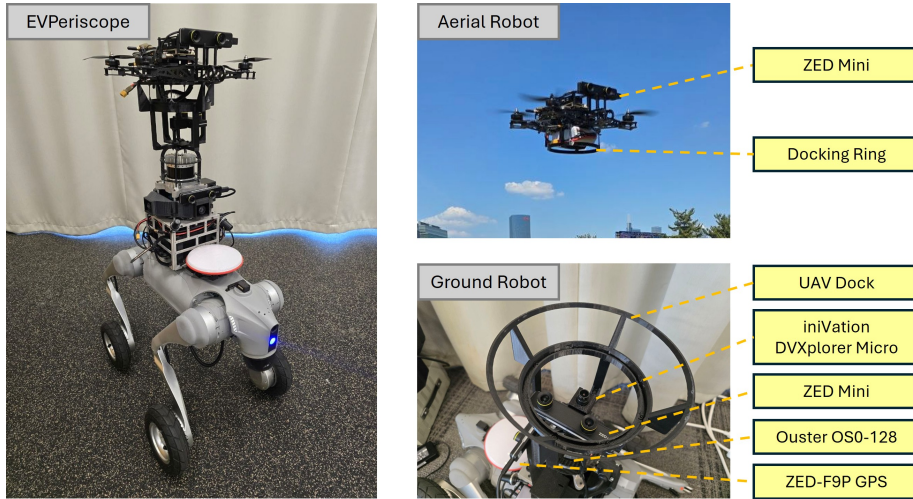


Fig. 3: System hardware configuration. EVPeriscope (left) consists of a custom-built quadrotor (upper right), and a Unitree Go2-W with a custom sensor stack (lower right). The core sensors of the system include the ZED Mini on the UAV for perception and state estimation, and the iniVation DVXplorer Micro event camera on the UGV for propeller tracking. A ZED Mini, Ouster OS0-128 and ZED-F9P GPS module on the UGV are also used for baseline comparisons. A custom docking system supports carrying and landing the UAV.

within 0.05 m of the target setpoint for at least 3 s before proceeding. When this condition is satisfied, the PX4 land mode is triggered.

3.5 Platform Details

An overview of the platform is shown in Fig. 3. The Kestrel UAV is a custom-built quadrotor with a 31.2 cm wheelbase and 7-inch propellers. The platform has a PX4 flight controller and is equipped with a forward-facing Stereolabs ZED Mini for state estimation. The sensors on the Unitree Go2-W UGV include an iniVation DVXplorer Micro event camera, upward-facing and forward-facing ZED Mini cameras, an Ouster OS0-128 LiDAR and a ZED-F9P GPS module. The event camera has a resolution of 640×480 and an FOV of approximately $90^\circ \times 70^\circ$. The UGV has an NVIDIA Jetson AGX Orin for compute. Communication between robots is supported by Rajant mesh radios. A custom-built docking system supports carrying and landing the UAV. The docking ring on the UAV is 14 cm in diameter. The docking pad on the UGV has a matching ring and is 30 cm in diameter at the other end.

The software stack runs on ROS 2 Humble over Zenoh middleware. All event processing, propeller tracking, PnP/EKF pose estimation, and relative-pose control are executed on the UGV's Jetson AGX Orin. The UAV computes depth and visual-inertial odometry onboard using the ZED Mini and Orin NX. The UAV transmits these perception and state-estimation outputs to the UGV, while the UGV sends velocity commands to the UAV flight controller.

Table 1: Quadrotor pose estimation error at different altitudes above the UGV. The position and attitude estimates from EVPeriscope are compared against the estimate state from the UAV onboard VIO. The RMSE and standard deviations of position and attitude errors over each altitude sequence are reported.

Altitude	Position (m) / Attitude (deg) Error					
	X	Y	Z	Roll	Pitch	Yaw
0.5	0.014 ± 0.006	0.027 ± 0.007	0.017 ± 0.005	1.975 ± 0.915	1.894 ± 1.137	1.342 ± 0.103
1.0	0.065 ± 0.029	0.086 ± 0.015	0.023 ± 0.014	2.250 ± 1.856	1.970 ± 1.522	1.611 ± 0.155
2.0	0.084 ± 0.015	0.125 ± 0.024	0.129 ± 0.019	2.744 ± 2.419	5.162 ± 3.645	1.944 ± 0.320
4.0	0.142 ± 0.040	0.434 ± 0.035	0.493 ± 0.032	4.248 ± 3.131	9.432 ± 9.344	4.752 ± 0.576
6.0	0.364 ± 0.084	0.546 ± 0.051	0.665 ± 0.055	7.260 ± 6.914	8.760 ± 6.537	8.500 ± 0.868

The ground robot runs the UAV detection and control modules on the Orin AGX at a fixed rate of 200 Hz — this includes communication with the UAV. The event stream is processed in 5 ms non-overlapping windows. Each iteration of our control loop takes on average 2.53 ms, with all the modules—propeller detection and tracking, UAV pose estimation and the controller—running synchronously.

4 Experiments

4.1 UAV Detection and Pose Estimation

We first compute the error of the estimated quadrotor pose at different distances from the event camera to understand the performance of our UAV detection and localization method.

Setup The ground robot is stationary and the UAV is commanded to a setpoint at different heights above the UGV. 1 minute of data is collected per sequence at various heights up to 6 m. We use the pose estimate from ZED VIO as the reference estimate of the UAV pose.

We also compare our approach to fiducial-based detection methods with conventional frame-based cameras and identify failure cases of such methods. We consider two scenarios: an RGB camera on the ground robot facing upwards to detect a fiducial marker mounted on the quadrotor, and a downward-facing RGB camera mounted on the quadrotor detecting a fiducial marker on the ground robot. We use a $0.1 \text{ m} \times 0.1 \text{ m}$ AprilTag marker that matches the footprint of the UAV docking ring.

Finally, we conduct an experiment to demonstrate operation in extreme low-light conditions. The same concept of active event generation with propellers is used in the dark by mounting LED propellers on the UAV. Since VIO is severely degraded in such illumination conditions, GPS is used for state estimation in the *night* experiment. This is the only experiment using GPS.

Results Table 1 summarizes the quadrotor pose estimation error at different altitudes. We present the Root Mean Square Error (RMSE) of the pose estimate for positions and

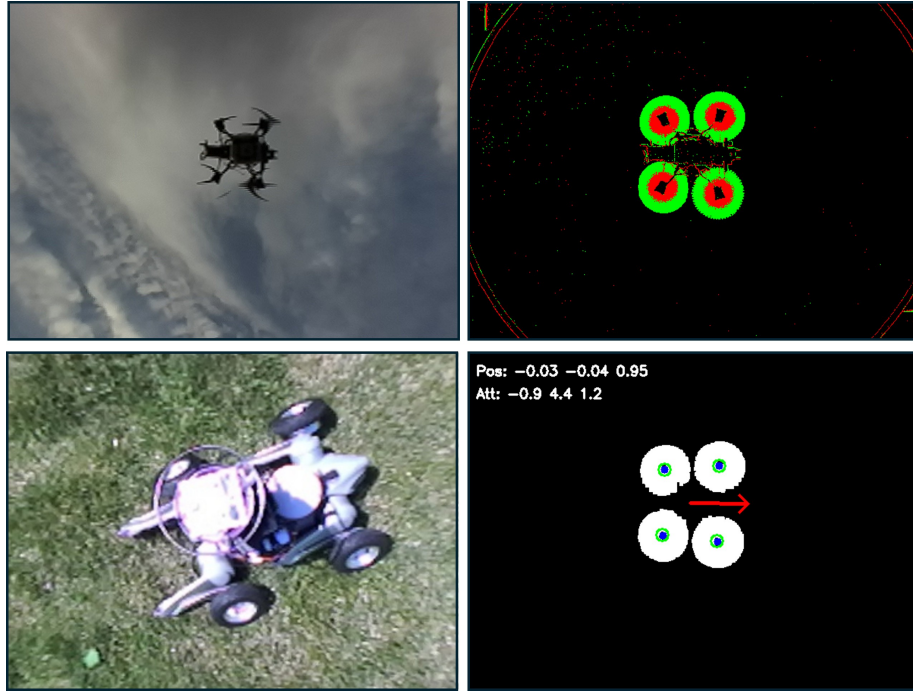


Fig. 4: Failure cases of UAV localization with conventional cameras. Even in overcast conditions, upward-facing imagery is often underexposed due to the bright backdrop of the sky (upper left). On the other hand, downward-facing cameras may struggle to handle the strong reflections of the fiducial markers (lower left). Conventional cameras likewise fail in low illumination conditions (middle). The harsh lighting conditions of outdoor environments make it challenging for fiducial marker detection using conventional frame-based cameras. In contrast, event-based cameras provide a strong signal on the propellers across a wide range of lighting conditions (upper right). The blue circles denote the estimated 2D propeller centroids, the green circles are the projected 3D centroids, and the red arrow indicates the estimated yaw of the UAV (lower right).

attitude along different axes along with their standard deviations. As expected, the farther the UAV from the camera, the higher the error in the pose estimation due to the resolution and field-of-view of the event camera. We show that we can localize the UAV with a reasonable margin of error for control—the positional error is less than 0.03 m when the quadrotor is near the camera and less than 0.7 m error at 6 m height. At larger distances, pose estimation accuracy is limited by the resolution and FOV of the event camera. In this work, we use a lens with a wider FOV to enable good tracking at small distances so as to support precise autonomous landing. Depending on the application, a narrower FOV may be preferred to improve tracking of the UAV at higher altitudes.

We highlight the challenges of frame-based detection of fiducial markers in Fig. 4. Conventional cameras lack sufficient dynamic range to balance exposure due to the sunlight, and struggle in darker conditions and farther distances as motion blur becomes more prominent. We ran the same set of experiments with the ZED Mini and a 100 mm × 100 mm AprilTag marker and found that marker detections were inconsistent at 0.5 m from

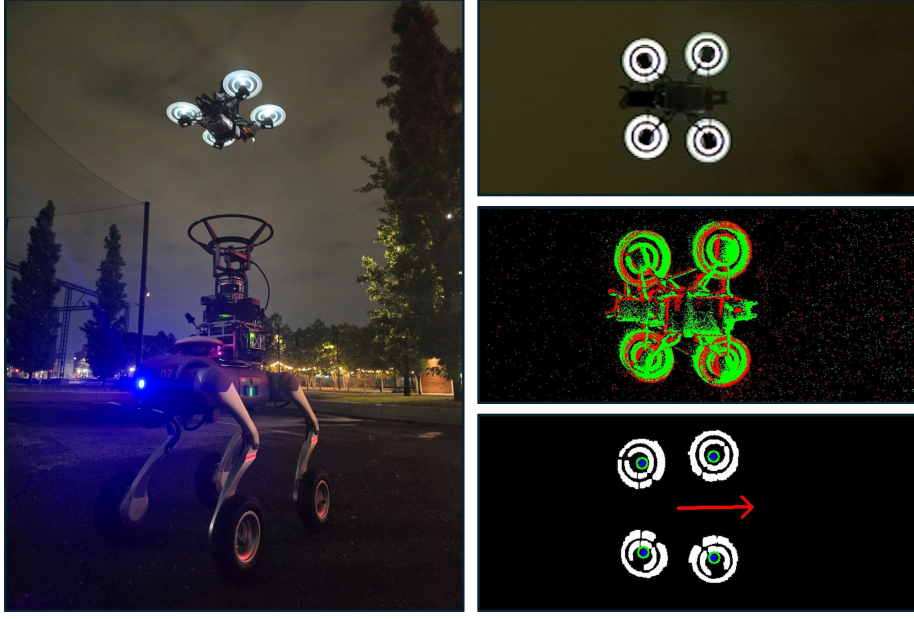


Fig. 5: Nighttime operation with EVPeriscope. The ground robot with the deployed aerial robot is shown on the left. The RGB camera image (upper right), event stream (middle right) and tracked propellers and estimated UAV pose (lower right) are visualized. With LED propellers to aid event generation, the framework is capable of propeller detection and tracking in low-light conditions.

the camera and failed beyond that. This is due to the fact that the platform size limits the size of the markers and harsh outdoor lighting conditions also lead to deterioration. Our approach can consistently detect the propellers and quadrotor across lighting conditions. This underscores the advantages of event-based cameras when operating out in the field under challenging conditions such as harsh sunlight.

Lastly, we show qualitative results of EVPeriscope operating at night in Fig. 5. With LED propellers, the aerial platform can still generate a strong signal in the event stream, enabling detection and tracking of the UAV with our method even in the dark. This demonstrates the potential of EVPeriscope for operation throughout the day and night regardless of lighting conditions.

4.2 Localization under Sensor Occlusion

We next show that our approach enables a ground robot to localize itself within dense foliage. By leveraging the extended perception provided by the quadrotor in the EVPeriscope framework, we show that the UGV is able to maintain accurate localization for navigation even when its own navigation sensors are occluded.

Setup We manually tele-operate the UGV through dense foliage in a field. We record the VIO state estimate from the quadrotor in EVPeriscope, the VIO estimate from the forward facing RGB camera on the UGV, the LiDAR-inertial odometry [26] (LIO) from the LiDAR on the UGV, and the GPS measurements of the UGV for ground truth.

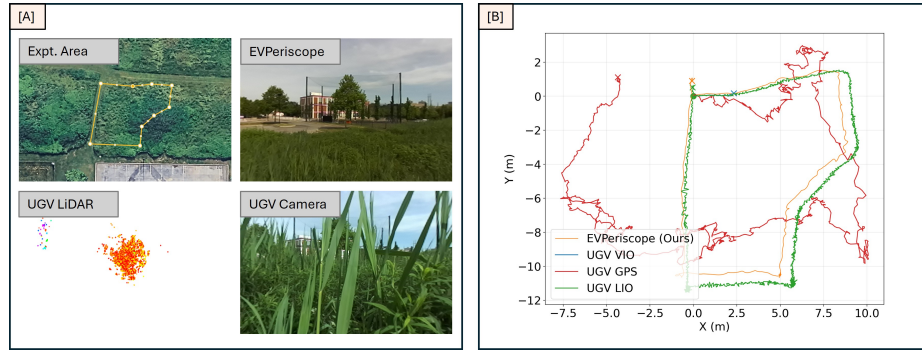


Fig. 6: Visualization of results for localization in dense foliage. [A] The ground robot is teleoperated to perform a loop through the field (top left). Sensor measurements of the periscope and ground robot are visualized. The UGV sensors are occluded by the tall grass whereas the UAV is unobstructed. [B] The estimated trajectories from each sensor are visualized. The start of each trajectory is denoted by $\circ$ and the end by $\times$. EVPeriscope is able to maintain a high vantage point above the foliage to enable accurate localization.

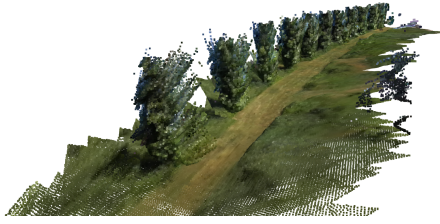
Results The results are presented in Fig. 6. The forward-facing ZED VIO completely fails due to occlusion of the sensor. Since the GPS antenna is not mounted high enough on the robot to be above the foliage, GPS measurements also exhibit a large amount of noise, highlighting the importance of alternative localization methods in such scenarios. Due to the wide field-of-view, high scan density and long range of the LiDAR sensor, the LIO is still able to capture global structure from faraway trees and lampposts and maintains a reasonable state estimate overall. However, there is significant noise along the trajectory which is detrimental to mapping and navigation. This is also dependent on how high the LiDAR can be mounted on the ground robot. Additionally, a majority of the LiDAR points are still occluded as shown in Fig. 6 and building a detailed map of the field with the LiDAR would not be possible. EVPeriscope maintains good perception and localization above the foliage for navigation.

4.3 Mapping with Extended Perception

We investigate the capability of EVPeriscope to enable mapping of areas of the environment that are typically beyond the field-of-view of UGV sensors by leveraging the elevated perspective of the UAV periscope.

Setup The system is tele-operated along a row of trees in the field with the objective of constructing a map of the environment. We build a map from the measurements from the UGV front-facing camera, the EVPeriscope UAV camera and the LiDAR on the UGV. The map constructed by each method is voxelized to 0.1 m and the occupied volume of each map is computed.

Results Results are presented in Fig. 7. EVPeriscope is able to map more of the environment compared to the UGV camera. This is particularly useful when parts of the



Method	Volume Mapped (m ³)
UGV LiDAR	421.82
UGV Camera	146.13
EVPeriscope	314.25

Fig. 7: Visualization of the map constructed by EVPeriscope (left) and the volume mapped by each method (right). The UGV camera is limited in its field-of-view due to being close to the ground. The UGV LiDAR, with its larger range and field-of-view, is provided as an upper bound. EVPeriscope is able to map more of the environment compared to the UGV camera with the elevation advantage of the UAV. This is crucial when the areas of interest are at a much higher elevation than the ground robot, such as when mapping trees in orchards or infrastructure inspection.

environment are beyond the vertical field-of-view of the UGV sensors, such as buildings and trees. EVPeriscope can leverage the elevation advantage provided by the UAV to map these areas of interest. This provides the benefits of a UAV for mapping while enabling longer missions by having the UGV carry the UAV between flights.

4.4 Closed-loop Navigation in Dense Foliage

We demonstrate the full EVPeriscope system with closed-loop navigation and obstacle avoidance in dense foliage.

Setup The experiment setup is visualized in Fig. 8. The system is given a waypoint behind the tree. To reach the waypoint, the system has to navigate through tall grass while avoiding the tree. The UAV is first tasked to take off from its dock on the UGV. Once the UAV is detected in the event stream, it is then transitioned to the tracking mode and positioned at 1.5m above the ground robot. Sensor measurements including RGB and depth images, and state estimation from the periscope are transmitted to the UGV in real time which are then fused into a map for planning and obstacle avoidance. Finally, navigation to the waypoint is initiated.

Depth images and corresponding poses are integrated into an occupancy map. For this experiment, we approximate the UAV-to-UGV transform as fixed at the commanded tracking setpoint. The relative pose estimated by EVPeriscope can instead be used to update this transform online. A global search with A^* on the occupancy map provides a coarse obstacle-free path, while a motion primitive-based planner is used to obtain dynamically feasible and collision-free trajectories (shown in green in Fig. 8). A trajectory tracker computes velocity commands which are sent to the robot control interface.

We also evaluate the tracking accuracy of the periscope while the system is navigating the environment. The experiment was conducted in wind conditions of 10–15 mph.

Results Qualitative results of the demonstration are visualized in Fig. 8. The system enables the robots to cooperatively navigate through the dense foliage while avoiding the

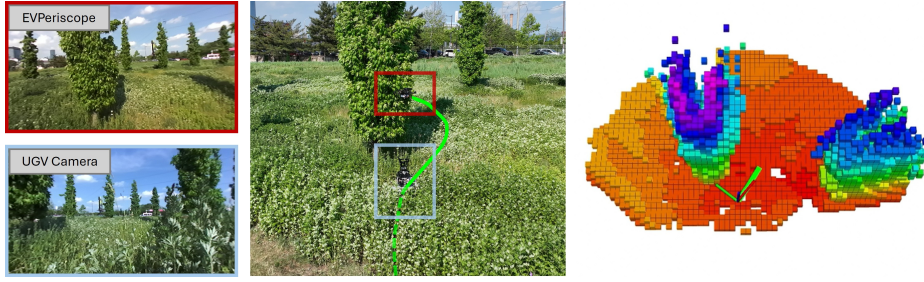


Fig. 8: EVPeriscope enables closed-loop navigation and obstacle avoidance in dense foliage where the UGV sensors are occluded. The sensor measurements of each robot are shown (left). The robots are highlighted in the scene (middle). Depth images and state estimates of the UAV are transmitted to the UGV, which performs mapping, planning and trajectory tracking with its onboard compute (right). The robot is indicated by the colored axes. The planned trajectory to avoid the obstacles is visualized as a green path.

tree. Along the trajectory, the closed-loop tracking RMSE is 0.125 m on position and 0.916° on yaw. This experiment demonstrates the capabilities of the EVPeriscope system, from robust quadrotor detection and tracking, to sharing of sensor measurements across robots for collaborative autonomy, with all modules running in real time on the platforms with onboard sensors and compute.

5 Discussion and Limitations

When the UGV sensors are occluded, the quality of the state estimation of the EVPeriscope system is limited by the localization sensor and algorithm on the UAV—visual-inertial odometry in this case. A larger UAV could be equipped with a LiDAR sensor for much more accurate localization, though payload constraints of the UGV have to be considered. While this work can filter out events generated by the quadrotor frame and background noise, it cannot handle high frequency noise such as from the propellers of another UAV. Future work could explore identifying and assigning propellers to unique UAVs to enable tracking and control of multiple robots.

When the ground robot’s sensors are unobstructed, the localization of both the UAV and UGV can be more tightly coupled, particularly since the UGV benefits from higher accuracy due to its LiDAR system. Although the UAV VIO-based state estimate may accumulate drift over time, the relative pose between the two platforms remains reliable. By fusing this relative pose information with the more accurate UGV state estimation, the system can leverage the aerial robot’s perception capabilities while minimizing the effects of its localization drift, making the most of both modalities.

6 Conclusion

We presented EVPeriscope, a marsupial UGV–UAV system that uses an upward-facing event camera on a ground robot to detect, localize, and control a quadrotor. The system enables tightly coupled coordination without external infrastructure. Through real-world

experiments, we showed that event-based propeller detection remains reliable under challenging outdoor lighting, supports localization and navigation when ground-level sensors are occluded by dense foliage, and expands mapping coverage by using the UAV as an elevated sensing platform. These results demonstrate the potential of this approach for robust heterogeneous robot teams in GPS-denied, cluttered, and perception-limited environments, with future work focused on tighter fusion of state estimation across robots, semantics-informed planning and long-range marsupial autonomy.

Bibliography

- [1] Rodriguez, M., Al-Kaff, A., Madridano, A., Martín, D., de la Escalera, A.: Wilderness search and rescue with heterogeneous multi-robot systems. In: 2020 International Conference on Unmanned Aircraft Systems (ICUAS). pp. 110–116 (2020)
- [2] Liao, C., Liu, Z.: Cooperative path planning of ground-air robots for distributed photovoltaic inspection. *Journal of Physics: Conference Series* 2658(1), 012015 (dec 2023)
- [3] Ju, C., Son, H.I.: Modeling and control of heterogeneous agricultural field robots based on ramadge–wonham theory. *IEEE Robotics and Automation Letters* 5(1), 48–55 (2020)
- [4] Olson, E.: Apriltag: A robust and flexible visual fiducial system. 2011 IEEE International Conference on Robotics and Automation pp. 3400–3407 (2011)
- [5] Sanket, N.J., Singh, C.D., Parameshwara, C.M., Fermüller, C., de Croon, G.C.H.E., Aloimonos, Y.: EVPropNet: Detecting drones by finding propellers for mid-air landing and following. In: *Proceedings of Robotics: Science and Systems (RSS)* (2021)
- [6] Lee, J., Choi, S.Y., Bretl, T.: The use of multi-scale fiducial markers to aid takeoff and landing navigation by rotorcraft. In: *AIAA SCITECH 2024 Forum*. p. 0338 (2024)
- [7] Springer, J., Kyas, M.: Autonomous drone landing with fiducial markers and a gimbal-mounted camera for active tracking. In: 2022 Sixth IEEE International Conference on Robotic Computing (IRC). pp. 243–247 (2022)
- [8] Eldeborg Lundin, A., Winzell, R., Hamrell, H., Gustafsson, D., Ovrén, H.: Drone detection using a low-power neuromorphic virtual tripwire. In: *European Conference on Computer Vision*. pp. 105–120. Springer (2024)
- [9] Chen, X., Xu, J., Ding, W., Wang, H., Luo, X., Duan, R., Chen, J., Wang, X., Liu, Y., Chen, X.: Count every rotation and every rotation counts: Exploring drone dynamics via propeller sensing. In: *Proceedings of the 2026 ACM/IEEE International Conference on Embedded Artificial Intelligence and Sensing Systems*. pp. 746–760 (2026)
- [10] Spetlik, R., Uhrová, T., Matas, J.: Efficient real-time quadcopter propeller detection and attribute estimation with high-resolution event camera. In: *Image Analysis: 23rd Scandinavian Conference, SCIA 2025, Reykjavik, Iceland, June 23–25, 2025, Proceedings, Part I*. p. 217–230. Springer-Verlag, Berlin, Heidelberg (2025)
- [11] Pawelczyk, M.L., Wojtyra, M.: Real world object detection dataset for quadcopter unmanned aerial vehicle detection. *IEEE Access* 8, 174394–174409 (2020)
- [12] Chen, N., Xiao, C., Dai, Y., He, S., Li, M., An, W.: Event-based tiny object detection: A benchmark dataset and baseline. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 7209–7218 (2025)
- [13] Magrini, G., Berlincioni, L., Becattini, F., Cultrera, L., Pala, P.: Drone detection with event cameras. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 4703–4714 (2025)

- [14] Terrence, S., Drouin, M.A., Picard, M., Dizeu, F.B.D., Orth, A., Gagné, G.: A virtual fence for drones: Efficiently detecting propeller blades with a DVXplorer event camera. In: *Proceedings of the International Conference on Neuromorphic Systems (ICONS)* (2022)
- [15] Wang, C., Li, W., Chen, W., Wang, Y., Liu, M., Gao, F.: Autonomous landing control of quadrotor uavs. In: *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*. pp. 276–280 (2015)
- [16] Xia, K., Shin, M., Chung, W., Kim, M., Lee, S., Son, H.: Landing a quadrotor uav on a moving platform with sway motion using robust control. *Control Engineering Practice* 128, 105288 (2022)
- [17] Araar, O., Aouf, N., Vitanov, I.: Vision based autonomous landing of multirotor uav on moving platform. *J. Intell. Robotics Syst.* 85(2), 369–384 (Feb 2017)
- [18] Zhao, Z., Han, P., Xu, Y., Xie, W., Zhang, W., Liang, K., Zeng, Q.: Vision-based autonomous landing control of a multi-rotor aerial vehicle on a moving platform with experimental validations. *IFAC-PapersOnLine* 55(3), 1–6 (2022), 16th IFAC Symposium on Large Scale Complex Systems: Theory and Applications LSS 2022
- [19] Polvara, R., Sharma, S., Wan, J., Manning, A., Sutton, R.: Vision-based autonomous landing of a quadrotor on the perturbed deck of an unmanned surface vehicle. *Drones* 2(2) (2018)
- [20] Raxit, S., Singh, S.B., Newaz, A.A.R.: Yolotag: Vision-based robust uav navigation with fiducial markers. In: *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*. pp. 311–316. IEEE (2024)
- [21] Springer, J., Kyas, M.: Autonomous drone landing with fiducial markers and a gimbal-mounted camera for active tracking. In: *2022 Sixth IEEE international conference on robotic computing (IRC)*. pp. 243–247. IEEE (2022)
- [22] Murphy, R.R., Ausmus, M., Bugajska, M., Ellis, T., Johnson, T., Kelley, N., Kiefer, J., Pollock, L.: Marsupial-like mobile robot societies. In: *Proceedings of the Third Annual Conference on Autonomous Agents*. pp. 364–365 (1999)
- [23] Hudson, N., Talbot, F., Cox, M., Williams, J., Hines, T., Pitt, A., Wood, B., Frousheger, D., Lo Surdo, K., Molnar, T., Steindl, R., Wildie, M., Sa, I., Kottege, N., Stepanas, K., Hernandez, E., Catt, G., Docherty, W., Tidd, B., Tam, B., Murrell, S., Bessell, M., Hanson, L., Tychsen-Smith, L., Suzuki, H., Overs, L., Kendoul, F., Wagner, G., Palmer, D., Milani, P., O’Brien, M., Jiang, S., Chen, S., Arkin, R.C.: Heterogeneous ground and air platforms, homogeneous sensing: Team CSIRO Data61’s approach to the DARPA subterranean challenge. *Field Robotics* 2, 595–636 (2022)
- [24] De Petris, P., Khattak, S., Dharmadhikari, M., Waibel, G., Nguyen, H., Montenegro, M., Khedekar, N., Alexis, K., Hutter, M.: Marsupial walking-and-flying robotic deployment for collaborative exploration of unknown environments. In: *2022 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. pp. 188–194. IEEE (2022)
- [25] Zacharia, A., Dharmadhikari, M., Alexis, K.: Collaborative exploration with a marsupial ground-aerial robot team through task-driven map compression. *IEEE Robotics and Automation Letters* (2025)
- [26] Malladi, M., Guadagnino, T., Lobefaro, L., Stachniss, C.: A robust approach for lidar-inertial odometry without sensor-specific modeling. *IEEE Robotics and Automation Letters* 11(6), 7420–7427 (2026)