

Reflection-Based Relative Localization for Cooperative UAV Teams Using Active Markers

Tim Lakemann¹, Daniel Bonilla Licea², Viktor Walter¹, Martin Saska¹

Abstract—Reflections of active markers in the environment are a common source of ambiguity in onboard visual relative localization. This work presents a novel approach that exploits these typically unwanted reflections for onboard relative localization in heterogeneous multi-Uncrewed Aerial Vehicle (UAV) teams operating over specular, near-planar surfaces such as calm water and flat indoor floors. Using only a camera, active markers, and a single-beam rangefinder, the method requires no prior knowledge of robot size, marker configuration, or surface reflectance properties, and propagates the uncertainty introduced by surface irregularities, including those of dynamic water surfaces. We further characterize, analytically and experimentally, how surface tilt biases the estimate, and identify the slope range over which the level-surface assumption holds. Indoor multi-UAV and outdoor experiments validate the localization accuracy, demonstrating reliable operation across varying lighting conditions and greater effective range than state-of-the-art vision-based active-marker methods, with higher accuracy beyond 30 m. The project website with the video is available under the following link: <https://timlakemann.github.io/reflection-based-relative-localization/>.

Index Terms—Aerial Systems: Perception and Autonomy, Localization, Multi-Robot Systems

I. INTRODUCTION

RELIABLE relative localization among team members is crucial for multi-robot systems to collaboratively execute tasks safely and reliably [1]. In outdoor environments, Global Navigation Satellite System (GNSS) often lacks the reliability needed for close-formation flight. Its signal accuracy and precision are particularly degraded in dense forests, natural landscapes, or urban canyons, and in some cases, it becomes entirely unavailable [1], [2]. External reference systems such as Real-Time Kinematic (RTK) antennas improve accuracy but constrain the operational area to the range of the ground station.

Onboard relative localization removes this dependence on external infrastructure. Decentralized Ultra-wideband (UWB) ranging provides centimeter-level distance accuracy [3], but requires active radio exchange between team members, making it dependent on communication and susceptible to radio frequency (RF) jamming. 3D Light Detection And Ranging

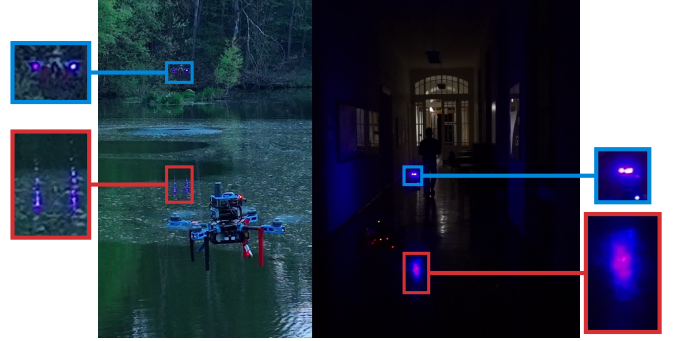


Fig. 1: Outdoor (left) and indoor dark (right) experiments: A UAV estimates the relative position by using surface reflections from active markers attached to a team member UAV. Blue boxes: the UAV with active markers attached. Red boxes: Surface reflections of the light emitted by the UAV.

(LiDAR) provides direct depth measurements robust to illumination, but it is heavy, power-intensive, and cannot associate a return with a specific team member [4]. Vision-based methods, in contrast, provide a cost-effective, compact, and lightweight solution that requires no communication between team members [5], [6]. While passive markers like AprilTags improve localization accuracy, they impose constraints on system dynamics, which is particularly critical for UAVs [7]. Active markers overcome limitations related to marker size, enable reliable detection under varying lighting conditions, and, through unique blinking sequences, link each detection to a known robot identity [8], [9]. Such systems, however, infer relative distance from the pixel separation between markers, and therefore require prior knowledge of the marker configuration on each UAV [9], [10]. As with marker-less or passive marker systems, the precision of this distance estimation decreases for Micro Aerial Vehicles (MAVs) with closely spaced markers. The limited physical spacing reduces pixel separation as distance increases, eventually falling below the resolution threshold required for reliable distance estimation [9].

This work addresses these limitations by introducing a reflection-based relative localization system for UAV teams operating above specular, near-planar surfaces, such as calm water in marine environments or polished floors in indoor facilities (Fig. 1).

Prior work has focused on mitigating surface reflections using polarization filters [11], [12] or structured light [13]. Reflection-based localization has also been explored in the RF spectrum [14], [15], [16] and acoustic domains [17], [18]. In [19], depth images are generated from water surface reflections. However, the approach requires a stationary camera and

Manuscript received: Aug. 05, 2026; Accepted Oct. 06, 2027.

This paper was recommended for publication by Editor Soon-Jo Chung upon evaluation of the Associate Editor and Reviewers' comments. This work was funded by the CTU grant no SGS26/077/OHK3/1T/13 and the Czech Science Foundation (GAČR) under research project no. 26-22419S.

¹ Tim Lakemann, Viktor Walter and Martin Saska are with the Department of Cybernetics, Czech Technical University in Prague, 14900 Prague, Czech Republic (e-mail: lakemtim@fel.cvut.cz).

² Daniel Bonilla Licea is with the Mohammed VI Polytechnic University in Morocco.

Digital Object Identifier (DOI): see top of this page.

40 seconds per frame, making it impractical for collaborative UAVs.

To the best of our knowledge, relative localization through surface reflections remains unexplored for multi-UAV teams in the literature. Unlike most vision-based relative localization methods, our approach exploits surface reflections to extend the operational range without requiring prior knowledge of the optical properties of the surface or of the size and marker configuration of the transmitting UAV.

By accounting for the spread of the reflection pattern caused by surface roughness (Fig. 2) and for sensor uncertainties, the approach remains reliable on surfaces that are not perfectly specular. It also removes the marker-spacing constraint discussed above, allowing the compact marker configurations typically required on MAVs. Our contribution in this work is threefold:

- 1) We propose a reflection-based relative localization approach that requires only a camera, active markers, and a single-beam rangefinder, with no prior knowledge of team member sizes, marker configurations, or optical surface properties.
- 2) We derive and experimentally validate the position bias induced by surface tilt, showing that it is asymmetric and that, for surfaces rising toward the transmitter, the estimates diverge at a finite range.
- 3) We validate the approach indoors and outdoors across heterogeneous platforms, demonstrating onboard operation, reliable performance under varying lighting, and improved accuracy over the state-of-the-art at ranges beyond 30 m.

II. REFLECTION-BASED LOCALIZATION METHOD

The proposed reflection-based relative localization method works effectively both indoors and outdoors, regardless of UAV size or prior knowledge of optical surface properties. Each UAV emits a predefined unique blinking sequence via its Light-emitting diodes (LEDs), generated following [20], where the sequence design, Hamming-distance separation, and resulting crosstalk and association-error behavior are analyzed in detail. The sequences are tracked and uniquely identified using the approach of [21], which was originally designed for low-exposure images in which each LED appears as a small, bright dot of only a few pixels. At higher exposure settings, required for reliable reflection detection, a single LED may be segmented into multiple detected regions. To address this, detections within a radius of r_c pixels are grouped via a clustering step, with the dominant cluster, i.e., the one with the highest number of extracted signals, treated as the direct LED source. If an additional cluster exists and is located vertically below the dominant cluster, it is selected as the reflection candidate. The position of each cluster is represented by its centroid, computed as the mean of all detections within the cluster. In the following, p_d denotes the centroid of the direct UAV-emitted light source and p_r the centroid of its reflection. When the blinking sequence cannot be extracted from the reflected region, p_r is instead identified geometrically as the detection located vertically below p_d in the image. Since

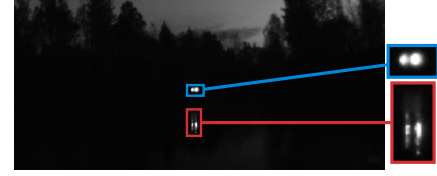


Fig. 2: Excerpt from an outdoor experiment showing active markers attached to the UAV (blue) and their diffuse reflections on the water surface (red), captured onboard the UAV.

the detected centroids are transformed into the gravity-aligned frame prior to any geometric reasoning, the vertical alignment criterion corresponds to physical verticality rather than image-axis alignment.

The proposed system was conceived for UAVs operating over specular, near-planar surfaces such as bodies of water with wave heights on the order of centimeters, or flat indoor floors, where spurious lateral reflectors are rare. When using LEDs as active markers, reflection intensity follows the Fresnel equations [22] and varies with surface smoothness, ranging from specular (calm water, polished floors) to diffuse (rough surfaces), with most indoor environments falling between these extremes [23, p. 108]. In practice, outdoor surfaces are further affected by environmental disturbances such as wind and propeller downwash, making explicit surface modeling intractable. While wall reflections are theoretically feasible, our experiments indicate that walls in urban indoor environments produce predominantly diffuse reflections, resulting in insufficient light intensity reaching the receiver (see Fig. 1 and top-left image in Fig. 4). Furthermore, reliable distance estimation from wall reflections would require knowledge of the wall orientation, necessitating a 2D LiDAR for vertically aligned walls or a 3D LiDAR for arbitrarily oriented surfaces. Incorporating such sensors would add significant weight and (computational) cost, contradicting the lightweight sensor philosophy of our approach. We therefore focus exclusively on ground and water reflections, for which a single dominant reflection is reliably obtained and the scalar distance to the surface is measured by a lightweight 1D rangefinder, such as a single-beam LiDAR or sonar, rather than the 2D or 3D LiDAR required to recover full surface orientation. The surface is consequently assumed level and near-planar. Local deviations from planarity are absorbed by the reflection cone model, and the position bias incurred by deviations from level is quantified in Sec. III-C. For simplicity, the following derivation considers a single transmitting UAV.

For a calibrated fisheye camera, the pixel coordinates of p_d and p_r can be converted into 3D unit bearing vectors using the function *cam2world*, which is available in various computer vision libraries, such as *OCamLib* [24]. In the following, the coordinate origin is defined at the body frame of the observer. The x -axis aligns with the horizontal center of the camera and is parallel to the surface, the y -axis is also parallel to the surface, and the z -axis points upwards (Fig. 3a). To account for the roll (ψ) and pitch (ϕ) angles of the observing aircraft, the 3D unit bearing vectors corresponding to p_d and p_r are

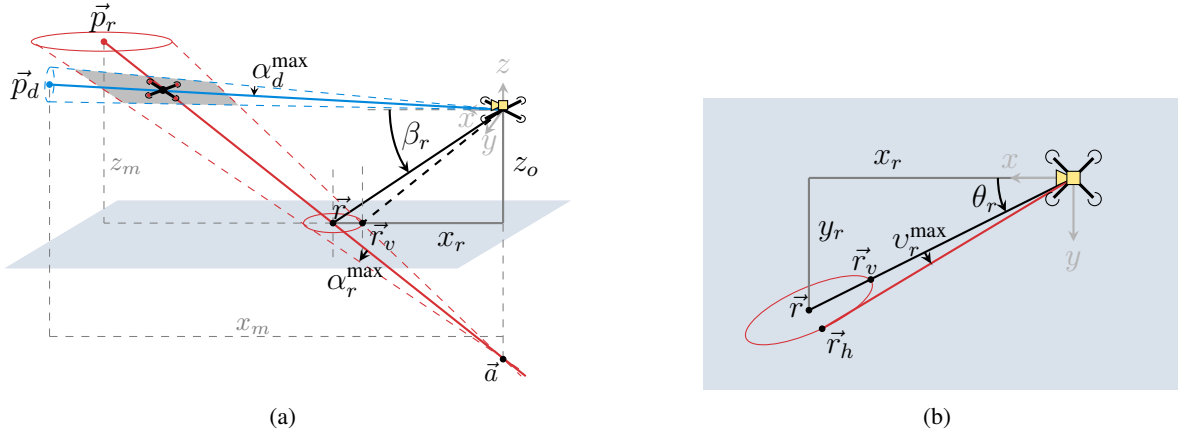


Fig. 3: One UAV equipped with a camera (yellow) constructs an elliptical cone based on the diffuse reflections of light emitted by the transmitting UAV (top left). (a) Side view of our approach, showing the potential location of the transmitting UAV determined by the intersection of two elliptical cones (gray area). (b) Top-down view of the observer UAV with the ellipse around the diffuse reflection.

transformed as follows:

$$\vec{b}_d = [\hat{x}_d, \hat{y}_d, \hat{z}_d]^\top = \mathbf{R}(-\psi)\mathbf{R}(-\phi)\text{cam2world}(p_d), \quad (1)$$

$$\vec{b}_r = [\hat{x}_r, \hat{y}_r, \hat{z}_r]^\top = \mathbf{R}(-\psi)\mathbf{R}(-\phi)\text{cam2world}(p_r), \quad (2)$$

where $\mathbf{R}(-\psi)$ and $\mathbf{R}(-\phi)$ are the rotation matrices that compensate for the roll and pitch of the observing UAV, respectively. While estimating the relative location from the intersection of the two bearing vectors is geometrically straightforward, surface irregularities and finite camera resolution introduce errors in both $\vec{b}_d$ and $\vec{b}_r$ that render this intersection ill-conditioned, producing large and unpredictable position errors. As shown in Fig. 3, the center of the reflection, denoted by $\vec{r}$, is retrieved using the two angles β_r and θ_r , which are defined as follows:

$$\beta_r = \arctan \frac{|\hat{z}_r|}{\hat{x}_r}, \quad (3)$$

$$\theta_r = \arctan \frac{\hat{y}_r}{\hat{x}_r}. \quad (4)$$

Thus, the center of the reflection, $\vec{r}$, is determined using equations (2) to (4), along with the height of the observer, z_o :

$$\begin{aligned} \vec{r} &= [r_x \ r_y \ r_z]^\top \\ &= \left[|z_o| \frac{\hat{x}_r}{|\hat{z}_r|} \quad |z_o| \frac{\hat{y}_r}{|\hat{z}_r|} \quad -z_o \right]^\top. \end{aligned} \quad (5)$$

As illustrated in Fig. 1 and 2, the light emitted by the UAV and its reflection on a surface appears elongated along the viewing axis. Two factors cause this elongation: the geometric projection of the reflected light at shallow incidence angles and the scattering induced by surface roughness. To represent the spread in the reflection, we construct an ellipse centered at p_r . The minor and major axes are defined by the farthest pixels from p_r in the vertical (p_v) and horizontal (p_h) directions, respectively. These pixels are selected based on their intensity exceeding a predefined threshold σ within a local bounding box around p_r . To retrieve the vectors representing the minor

and major axis of the ellipse, equations (2) to (4) can be applied to p_v and p_h yielding:

$$\vec{r}_v = \left[|z_o| \frac{\hat{x}_v}{|\hat{z}_v|} \quad |z_o| \frac{\hat{y}_v}{|\hat{z}_v|} \quad -z_o \right]^\top. \quad (6)$$

$$\vec{r}_h = \left[|z_o| \frac{\hat{x}_h}{|\hat{z}_h|} \quad |z_o| \frac{\hat{y}_h}{|\hat{z}_h|} \quad -z_o \right]^\top. \quad (7)$$

Using equations (5) to (7), we construct an elliptical cone (red in Fig. 3a). The apex of the cone is defined by the vector $\vec{a} = [0 \ 0 \ -2z_o]^\top$, and the half-apex angle of the cone along the vertical axis is $\alpha_r^{\max} = \beta_v - \beta_r$, which corresponds to the major axis of the ellipse defined by the distance between $\vec{r}$ and $\vec{r}_v$ (Fig. 3a). As shown in Fig. 3b, the horizontal spread of the cone along the horizontal axis (corresponding to the minor axis of the ellipse) defines the half-apex angle:

$$v_r^{\max} = \arctan \frac{\|\vec{r}_h - \vec{r}\|_{xy}}{\|\vec{r} - \vec{a}\|_{xy}}, \quad (8)$$

in the x - y plane. Using a predefined maximum height (z_m), we define the end point of the elliptical cone by the vector:

$$\vec{p}_r = \begin{bmatrix} r_x + z_m \frac{\hat{x}_r}{|\hat{z}_r|} \\ r_y + z_m \frac{\hat{y}_r}{|\hat{z}_r|} \\ z_m \end{bmatrix}. \quad (9)$$

With equations (3) to (9), the following parametric equation defines a finite convex elliptical cone:

$$\begin{aligned} \vec{c}_r(\gamma_r, \alpha_r, v_r, l_r) &= \vec{a} + l_r \vec{d}_r + l_r \vec{n}_1 \tan \alpha_r \cos \gamma_r \\ &\quad + l_r \vec{n}_2 \tan v_r \sin \gamma_r, \end{aligned} \quad (10)$$

where:

$$\begin{aligned} \gamma_r &\in [0, 2\pi], & \alpha_r &\in [0, \alpha_r^{\max}], & v_r &\in [0, v_r^{\max}], \\ l_r &\in [0, \|\vec{p}_r - \vec{a}\|], & \vec{d}_r &= \frac{\vec{p}_r - \vec{a}}{\|\vec{p}_r - \vec{a}\|}, \end{aligned} \quad (11)$$

and $\vec{n}_1$ is a unit vector perpendicular to $\vec{d}_r$, while $\vec{n}_2$ is a unit vector perpendicular to both $\vec{d}_r$ and $\vec{n}_1$.

To address uncertainties arising from limited camera resolution and detection inaccuracies, we define two angles, $\alpha_d^{\max}$ and $v_d^{\max}$, which represent the vertical and horizontal uncertainties around $\vec{b}_d$, respectively. These uncertainties are used to create a finite convex elliptical cone around $\vec{b}_d$, similar to the cone representation of the reflection vector and its associated uncertainty in (10). This cone is parameterized by:

$$\vec{c}_d(\gamma_d, \alpha_d, v_d, l_d) = l_d \vec{d}_d + l_d \vec{n}_3 \tan \alpha_d \cos \gamma_d + l_d \vec{n}_4 \tan v_d \sin \gamma_d, \quad (12)$$

where:

$$\begin{aligned} \gamma_d &\in [0, 2\pi], & \alpha_d &\in [0, \alpha_d^{\max}], & v_d &\in [0, v_d^{\max}], \\ l_d &\in [0, \|\vec{p}_d\|], & \vec{d}_d &= \frac{\vec{p}_d}{\|\vec{p}_d\|}, \end{aligned} \quad (13)$$

and $\vec{n}_3$ is perpendicular to $\vec{b}_d$, and $\vec{n}_4$ is perpendicular to both $\vec{n}_3$ and $\vec{b}_d$. The vector denoting the end point of the apex axis, depends on the predefined maximum extent in the x -direction (x_m) and is expressed as:

$$\vec{p}_d = \begin{bmatrix} x_m & x_m \frac{\hat{y}_d}{\hat{x}_d} & x_m \frac{\hat{z}_d}{\hat{x}_d} \end{bmatrix}^\top. \quad (14)$$

The intersection area of the two cones, defined by equations (10) and (12), represents possible locations of the transmitting UAV, resulting in a system of non-linear equations. To estimate the location of the transmitting UAV we employ a random sampling-based approach: we generate a set of vectors $\mathcal{X} = \{\vec{p}_i \sim U|i = 1, \dots, n\}$ whose elements verify equation (10), and therefore ensuring they lie within the reflection cone. The vectors in $\mathcal{X}$ are generated by sampling the parametric variables from uniform distributions:

$$\begin{aligned} \gamma'_r &\sim U(0, 2\pi), & \alpha'_r &\sim U(0, \alpha_r^{\max}), \\ v'_r &\sim U(0, v_r^{\max}), & l'_r &\sim U(l_r^{\min}, \|\vec{p}_r - \vec{a}\|), \end{aligned} \quad (15)$$

where $l_r^{\min} = \|\vec{a} - \vec{r}\|$ ensures that only physically feasible points are sampled, i.e., points located above the surface and within the possible intersection volume. To determine whether $\vec{p}_i \in \mathcal{X}$ lies within the cone defined by equation (12) and shown in blue in Fig. 3, we first compute its projection onto the cone axis:

$$\vec{p}_{i||} = \text{proj}_{\vec{d}_d} \vec{p}_i = (\vec{p}_i \cdot \vec{d}_d) \vec{d}_d, \quad (16)$$

with $\cdot$ denoting the scalar product, the two inequalities must hold:

$$\frac{(\vec{p}_i - \vec{p}_{i||})_y^2}{(\|\vec{p}_{i||}\| \tan v_d^{\max})^2} + \frac{(\vec{p}_i - \vec{p}_{i||})_z^2}{(\|\vec{p}_{i||}\| \tan \alpha_d^{\max})^2} \leq 1, \quad (17)$$

$$0 \leq \|\vec{p}_{i||}\| \leq l_d. \quad (18)$$

Inequality (17) ensures that the point lies within the elliptical cross-section of the cone, while inequality (18) ensures that it lies within the extent of the cone along its axis. If a point in $\mathcal{X}$ satisfies both conditions given by equations (17) and (18), it is included in the inlier set $\mathcal{X}_d$. The points in $\mathcal{X}_d$ are then used as observations for a particle filter with n_p particles, which propagates according to a constant-velocity

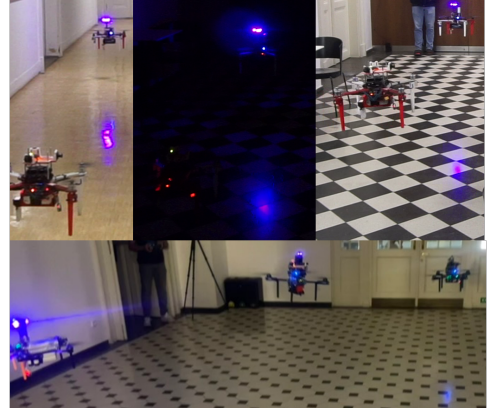


Fig. 4: Indoor experiments on different floor types (PVC, top left; tiles, top right) and under dark conditions (top center). Bottom: heterogeneous multi-UAV experiment.

model and updates particle weights using a robust Cauchy kernel. The particles are initialized using a uniform distribution over the 3D space observable by the camera. Resampling is triggered when the Effective Sample Size (ESS) falls below a predefined threshold, set to half the total number of particles. The resulting weighted particle cloud is then used to compute the final position estimate of the transmitting UAV.

III. EXPERIMENTS

We evaluated the approach in both indoor and outdoor experiments, comparing results against the state-of-the-art vision-based active-marker method. Unless stated otherwise, processing was performed post-flight on a local machine (*Intel i7-8550U*, 1.8 GHz). Onboard execution is demonstrated in the multi-UAV experiment (Sec. III-A2). Ground truth was provided by UWB indoors and RTK outdoors. In the following, we distinguish between two platform types. The UAV refers to platforms based on the *F450* or *Holybro X500* frame [25], equipped with an *Intel NUC 10 i7FNK* (6 cores, up to 4.7 GHz), while the MAV refers to a custom 27 cm platform equipped with a *Raspberry Pi 5* (4 cores, up to 2.4 GHz, 8 GB RAM). Both platforms carried a camera with a fisheye lens and Ultraviolet (UV) bandpass filter. The UAVs used a 752×480 BlueFOX camera operating at 10 Hz to 30 Hz, while the MAV carried an ArduCam operating at 640×400 and 10 Hz. While a downward-facing LiDAR-based rangefinder was used for height estimation indoors, RTK was employed outdoors. Both sensors relayed measurements to the onboard PC via MAVLink, with MAVROS providing clock synchronization and interpolating height measurements to camera timestamps. For outdoor experiments, we mounted UV-LEDs on the arms of the transmitter UAV, with all LEDs emitting an

Experiment	n	n_p	$\alpha_r^{\max}$ [°]	$\alpha_d^{\max}$ [°]	$v_d^{\max}$ [°]	x_m [m]	z_m [m]	σ	t_{exp} [ms]
Indoor 1–4	500	1500	3	3	3	15	2.5	30	20
Multi-UAV	500	1500	3	3	3	10	2.5	30	10
Outdoor 1–3	1000	1000	4	3	3	30–40	10	40	2–3

TABLE I: Parameters for indoor and outdoor experiments.

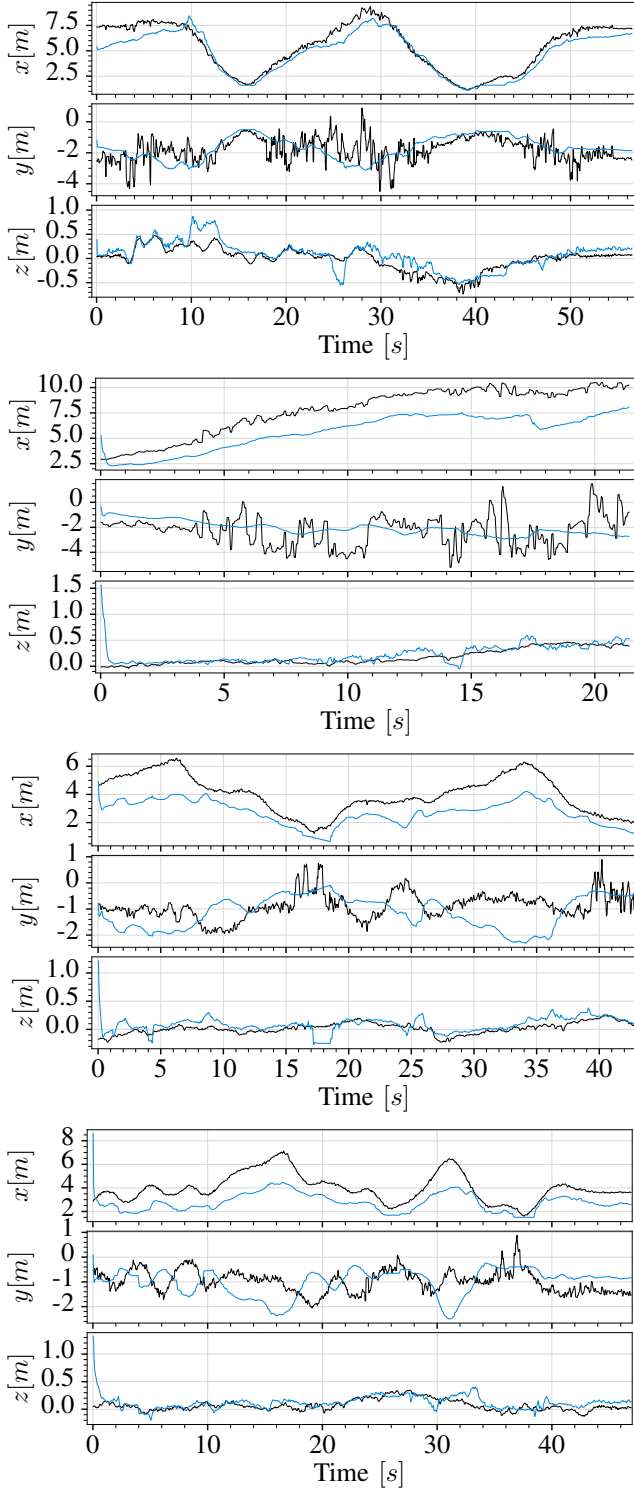


Fig. 5: Relative position estimates obtained using our approach (blue) compared to the relative position ground truth from UWB (black) during indoor experiments 1–4 (from top to bottom).

identical signal, consistent with the UltraViolet Direction And Ranging (*UVDAR*) setup in [9]. During the indoor and outdoor experiments the observing UAV was hovering with roll and pitch angles below 0.06 rad, so the camera frame was used as the gravity-aligned frame. Tab. I summarizes the parameters

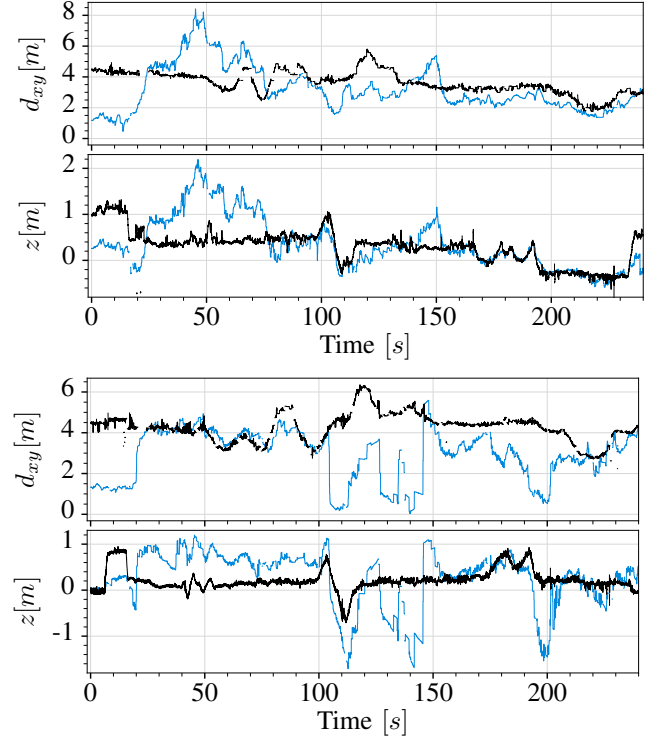


Fig. 6: Relative position estimates obtained using our approach (blue) and the ground truth from UWB (black) for the UAV (top) and MAV (bottom) during the multi-UAV experiment.

used in the experiments. Exposure time settings were adapted depending on the environment (indoor vs. outdoor), with the local binary threshold (σ) adjusted accordingly. Additionally, the number of samples in the reflection cone n was reduced in the multi-UAV experiment, to ensure onboard deployability on the MAV.

A. Indoor

Indoor experiments were conducted in two environments: a standard office hallway with a Polyvinyl chloride (PVC) floor and a corridor with a tiled surface (Fig. 4). To highlight platform independency, a single circular LED array was mounted on top of each UAV (Fig. 4). The *UVDAR* system could not be evaluated in this setup, as the LED array on the transmitting UAV had minimal spacing between LEDs, while the *UVDAR* system requires at least two spatially separated markers. Height estimation on the observer was performed using a downward-facing LiDAR-based rangefinder. The experiments are divided into two-UAV and multi-UAV configurations.

	1. PVC (d)	2. PVC (n)	3. Tiles (n)	4. Tiles (n)
MAE(x) [m]	1.29 ± 0.67	2.12 ± 0.70	1.30 ± 0.61	1.22 ± 0.60
MAE(y) [m]	0.65 ± 0.52	1.09 ± 0.74	0.68 ± 0.42	0.61 ± 0.41
MAE(z) [m]	0.12 ± 0.09	0.08 ± 0.12	0.10 ± 0.09	0.08 ± 0.09

TABLE II: Mean Absolute Error (MAE) in the x -, y -, and z -axes for indoor experiments conducted under daylight (d) and nighttime (n) conditions on different floor types. Values are reported as MAE $\pm$ one standard deviation.



Fig. 7: Outdoor experiment: (Left) Observing UAV (white), transmitting UAV with active markers (blue), and surface reflections (red). (Right) Onboard UV-filtered image showing the detected marker centroid (blue) and reflection ellipse (red).

1) *Two-UAV Experiment*: Fig. 5 shows the resulting relative position estimates compared to UWB ground truth. Tab. II summarizes the Mean Absolute Errors (MAEs) and standard deviations per axis across all experiments. Consistently, the largest MAE was observed along the x -axis (optical axis), which is expected as depth estimation from a monocular camera is inherently less constrained than lateral estimation. The method performed reliably across both floor types and under both daylight and nighttime conditions, demonstrating robustness to varying surface properties and lighting.

2) *Multi-UAV Experiment*: The system ran onboard all three UAV platforms, including the MAV equipped with a *Raspberry Pi 5*, demonstrating deployability on low-computational-power hardware. The clustering radius was set to $r_c = 10$ pixels to account for the higher exposure settings required for reflection detection. Fig. 6 shows the estimated relative position, decomposed into the planar Euclidean distance (d_{xy}) and vertical (z) components, from one UAV to another UAV and to one MAV. The method achieves an MAE of 1.50 ± 1.44 m in d_{xy} and 0.49 ± 0.36 m in z for the UAV, and 1.25 ± 0.93 m in d_{xy} and 0.36 ± 0.37 m in z for the MAV. The system reliably estimated the relative position of the UAV and MAV outside the interference interval from 105 s to 155 s, during which the MAV was deliberately positioned in front of the observer to induce signal interference. Full position recovery requires both a reliable blinking-sequence extraction and the direct bearing $\vec{p}_d$ under Line of Sight (LoS), both of which degrade in this configuration. Relative bearing alone remains obtainable but does not determine range. The per-timestamp computational complexity is dominated by the cone sampling and particle filter update, yielding $\mathcal{O}(n \cdot n_p)$ in the worst case, where n is the number of samples drawn in the reflection cone and n_p the number of particles. In practice, the cone sampling accounts for approximately 80 % of the per-frame latency, while the particle filter update remains below 3 ms. The median estimator latency was 78 ms on the *Intel NUC* and 97 ms on the *Raspberry Pi 5*, exceeding the 100 ms camera period for 19.6 % and 45.4 % of frames, respectively. As the estimator always processes the newest frame, an over-run lowers the update rate rather than accumulating latency, yielding mean rates of 8.4 Hz and 6.8 Hz at the accuracy reported above.

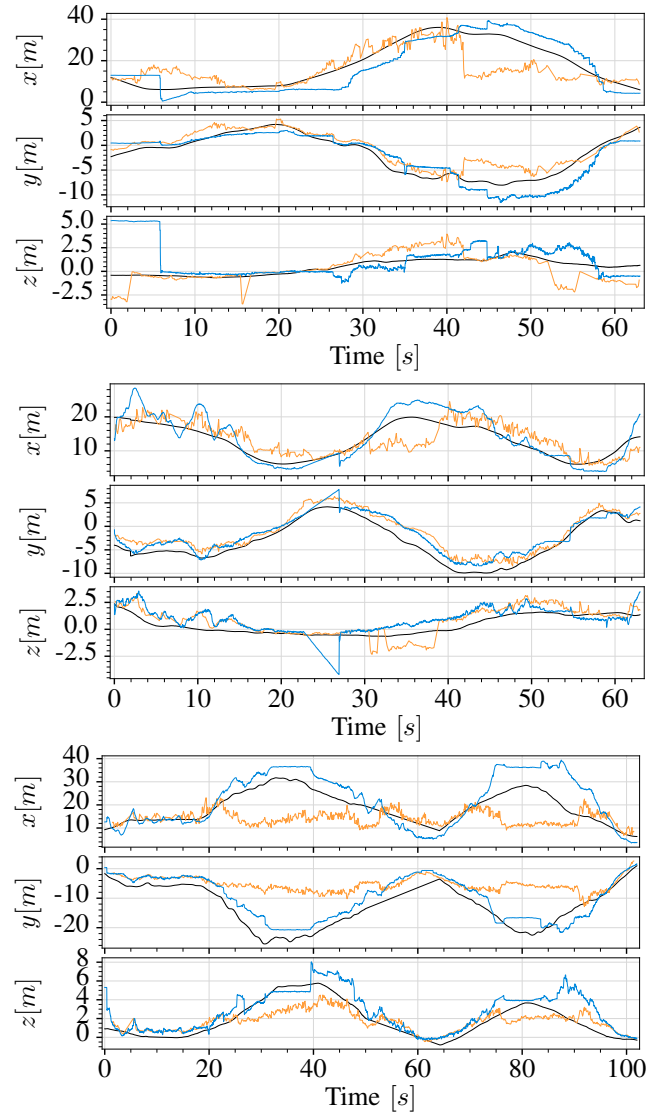


Fig. 8: Relative position estimates obtained using our approach (blue), UVDAR (orange), and the ground truth from RTK (black) for outdoor experiments 1–3 (from top to bottom).

B. Outdoor

Fig. 7 provides an overview of the outdoor experimental setup. Multiple outdoor experiments with varying exposure time settings were conducted (Tab. I). For example, in the experiment shown in Fig. 2, automatic exposure and gain control were used. We observed that with a maximum exposure time of 3 ms, the camera maintained a stable frame rate (essential for reliable signal extraction) while reflections remained clearly visible at distances exceeding 20 m on the water surface.

Fig. 8 shows the three most relevant outdoor flights, comparing the relative positions obtained from our method, UVDAR, and the ground truth. Our reflection-based approach maintained high accuracy at distances beyond 30 m, whereas UVDAR struggled to provide reliable estimates past 20 m. This highlights the advantage of our approach, which operates without prior knowledge of the UAV size or the configuration

		X			Y			Z		
		MAE[m]	<i>p</i> -value	95 % CI	MAE[m]	<i>p</i> -value	95 % CI	MAE[m]	<i>p</i> -value	95 % CI
1. Outdoor	Ours	3.50 ± 1.86	1.9×10^{-8}	[-2.62, -1.61]	1.18 ± 0.80	0.015	[-0.26, -0.07]	1.04 ± 1.57	0.917	[0.15, 0.42]
	UVDAR	5.58 ± 5.77			1.37 ± 1.09			0.78 ± 0.83		
2. Outdoor	Ours	2.42 ± 1.50	2.0×10^{-5}	[-0.75, -0.29]	1.73 ± 0.95	0.071	[-0.16, 0.02]	0.70 ± 0.47	0.009	[-0.01, 0.07]
	UVDAR	2.97 ± 2.14			1.76 ± 0.97			0.67 ± 0.44		
3. Outdoor	Ours	5.01 ± 4.15	2.2×10^{-11}	[-1.95, -1.18]	3.29 ± 1.58	0.000	[-4.89, -4.19]	0.98 ± 0.85	0.793	[-0.04, 0.11]
	UVDAR	7.00 ± 6.03			7.96 ± 5.60			0.97 ± 0.67		

TABLE III: Mean Absolute Error (MAE) and one standard deviation along the x -, y -, and z -axes, together with the *Wilcoxon* p -values and 95 % Confidence Interval (CI) of the paired differences, comparing our method with UVDAR in outdoor experiments.

of its light sources.

Tab. III summarizes the MAEs and standard deviations for each axis across all experiments, along with the *Wilcoxon* p -values and 95 % CI of the paired differences. Along the z -axis, UVDAR achieved slightly lower MAE values (1 % to 25 %). However, the differences were not statistically significant. Along the y -axis, our method reduced the MAE by 13.9 % and 58.7 % in experiments 1 and 3 (p -value < 0.05, 95 % CI fully below zero), whereas no statistically significant difference was observed in experiment 2. Along the x -axis (optical axis), our method reduced the MAE by 18.5–37.3 % compared to UVDAR, with the difference being statistically significant (p -value < 0.05, 95 % CI fully below zero).

C. Robustness to Model Assumptions

We evaluate robustness to two violations of the idealized model. The first is error in the height estimation by the observer. The second is tilt of the reflective surface. Height error is characterized on the outdoor water dataset, where barometric drift is a realistic failure mode. Surface tilt is characterized indoors, the setting where floor slope is the relevant deviation from the level-surface assumption.

To assess sensitivity to height error (e.g., from barometric sensors), we introduced systematic biases of up to $\pm 15\%$, representing a worst-case scenario. Tab. IV reports the outdoor results, which exhibit the largest absolute errors across our experiments, yet show only minor variation in localization accuracy across all evaluated scenarios. During the same flights, an ultrasonic rangefinder on the transmitting UAV deviated from the RTK height by 4 cm RMS, which is below 4 % of the flight height of the observer and well within this range.

To assess sensitivity to surface tilt, we analyzed the localization error for tilt angles τ up to $\pm 15^\circ$. We physically tilted the observer and bypassed the attitude compensation for the tilt axis, which is equivalent to a level observer over a surface tilted by τ . We define $\tau < 0$ as a tilt that raises the surface

	x		y		z	
	-15 %	+15 %	-15 %	+15 %	-15 %	+15 %
1. Outdoor	+0.33	-0.23	-0.03	-0.01	+0.02	-0.04
2. Outdoor	+0.07	+0.14	+0.39	-0.16	-0.05	+0.09
3. Outdoor	+0.92	+1.22	+0.10	-0.55	+0.30	+0.15

TABLE IV: Sensitivity analysis: change in MAE [m] under systematic bias in the height measurement of the observer UAV, relative to the unbiased results in Tab. III.

toward the transmitter and $\tau > 0$ as a tilt away from it. Fig. 9 shows the theoretical error against range alongside the measured results, which follow the predicted trend. For negative τ , a singularity occurs at $d_{sing} = h / \tan |\tau|$, at which the estimate diverges. For positive τ , the range is underestimated but remains bounded. The level-surface assumption thus holds for nearly level surfaces such as calm water, while at $|\tau| = 5^\circ$ the error exceeds that of the level-surface experiments beyond about 8 m (Fig. 9).

IV. DISCUSSION

Tab. V compares our approach against representative state-of-the-art systems. The LiDAR-based approach in [4] requires substantially higher onboard compute for point cloud processing than the camera-based methods, and its near-infrared pulses are largely absorbed by water, limiting its applicability over the outdoor water surfaces targeted in this work. The compared methods also differ in what must be known a priori. UVDAR assumes a known marker geometry, restricting deployment to platforms large enough to carry sufficiently spaced markers. CREPES achieves accurate ranging but requires UWB transceivers on all team members, adding a radio dependency. The CNN-based approach depends on a training dataset and on consistent lighting conditions. Unlike onboard stereo cameras, whose baseline is limited by the size of the UAV [26], the observer and its mirror image form a vertical baseline of $2h$. Our method requires only the range to the reflective surface, which is directly measurable onboard, and remains independent of transmitter size and marker config-

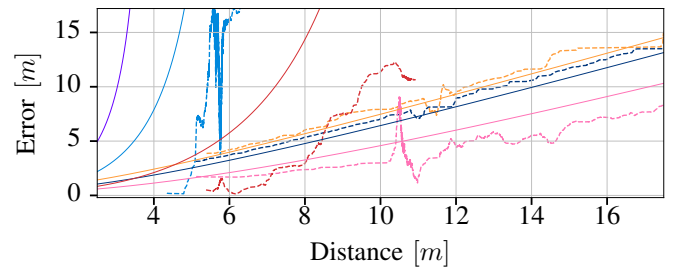


Fig. 9: Euclidean localization error vs. range for surface tilt angles 15° , 10° , 5° , -5° , -10° , -15° . Solid lines show the theoretical prediction, dashed lines the measured moving average. No measurement is shown for $\tau = -15^\circ$, where the geometry is singular within the evaluated range and the method produced no output.

	Ours	UVDAR [9]	CREPES [3]	Swarm-LIO2 [4]	CNN [6]
Hardware	Cam, LED, Rng	Cam, LED	Cam, LED, UWB, IMU	LiDAR, IMU	Cam, GPU
Mass / power [g / W] ¹	62 / 6.5	40 / 5.9	130* / $\leq 4^*$	265 / 6.5	200* / 17*
Output rate [Hz]	10	60	100	30	5
Error @ range [m] ²	MAE _x 2.4–5.0 @ ≤ 40	MAE _x 3.0–7.0 @ ≤ 40	P25–P75 0.10–0.16 @ ≤ 27.5	std. 0.05	P25–P75 0.3–0.7 @ 6
Approx. cost [\$]	570	450	500*	700*	500*
Direct ID	✓	✓	✓	✗	✗
RF-free	✓	✓	✗	✓	✓
Lighting-robust	✓	✓	●	✓	✗

TABLE V: Comparison of onboard relative localization approaches. Symbols: ✓ yes, ● partial, ✗ no. ¹Sensing hardware only, onboard compute excluded. ²Position error as reported in the original publications, following the convention of each work. MAE_x: mean absolute error along the optical axis, P25–P75: interquartile range, std.: swarm-state consistency standard deviation reported by the authors, as mutual-distance ground truth was unavailable. The metrics are not directly comparable. *Estimated from datasheets or manufacturer specifications.

uration, while enabling direct optical identification without reliance on radio communication.

V. CONCLUSION

This work presents an approach for onboard relative localization in UAV teams. Our method estimates the relative position between UAVs by exploiting typically unwanted reflections of active localization markers attached to team members. By accounting for uncertainties introduced by camera resolution and surface roughness, the proposed method operates without prior knowledge of team member sizes or optical surface properties. We validated our approach across indoor and outdoor experiments, demonstrating reliable performance under varying lighting conditions and across different floor materials. We further characterized the position bias induced by surface tilt. Outdoor experiments further showed that our method outperforms the state-of-the-art active-marker system, particularly at greater distances (30 m), reducing the MAE by up to 37.3% along the optical axis.

REFERENCES

- [1] T. H. Nguyen and L. Xie, "Relative transformation estimation based on fusion of odometry and uwb ranging data," *IEEE Transactions on Robotics*, vol. 39, no. 4, pp. 2861–2877, 2023.
- [2] Z. Zhang, H. Yuan, X. He, B. Li, and J. Geng, "Best Integer Equivariant Estimation With Quality Control in GNSS RTK for Canyon Environments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 4, pp. 4105–4117, 2023-08.
- [3] Z. Xun *et al.*, "CREPES: Cooperative Relative Pose Estimation System," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023-10, pp. 5274–5281.
- [4] F. Zhu *et al.*, "Swarm-LIO2: Decentralized Efficient LiDAR-Inertial Odometry for Aerial Swarm Systems," *IEEE Transactions on Robotics*, vol. 41, pp. 960–981, 2025.
- [5] H. Xu, L. Wang, Y. Zhang, K. Qiu, and S. Shen, "Decentralized Visual-Inertial-UWB Fusion for Relative State Estimation of Aerial Swarm," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020-05, pp. 8776–8782.
- [6] F. Schilling, F. Schiano, and D. Floreano, "Vision-Based Drone Flocking in Outdoor Environments," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2954–2961, 2021-04.
- [7] E. Olson, "Apriltag: A robust and flexible visual fiducial system," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 3400–3407.
- [8] H. Stuckey, L. Escamilla, L. R. Garcia Carrillo, and W. Tang, "Real-Time Optical Localization and Tracking of UAV Using Ellipse Detection," *IEEE Embedded Systems Letters*, vol. 16, no. 1, pp. 1–4, 2024-03.
- [9] V. Walter, N. Staub, A. Franchi, and M. Saska, "UVDAR System for Visual Relative Localization With Application to Leader-Follower Formations of Multirotor UAVs," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2637–2644, 2019-07.
- [10] M. Faessler, E. Mueggler, K. Schwabe, and D. Scaramuzza, "A monocular pose estimation system based on infrared leds," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 907–913.
- [11] K. Berger, R. Voorhies, and L. H. Matthies, "Depth from stereo polarization in specular scenes for urban robotics," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 1966–1973.
- [12] Y. Li, S. Lin, H. Lu, S. B. Kang, and H.-Y. Shum, "Multibaseline stereo in the presence of specular reflections," in *2002 International Conference on Pattern Recognition*, vol. 3, 2002, pp. 573–576.
- [13] X. Huang *et al.*, "Polarization structured light 3d depth image sensor for scenes with reflective surfaces," *Nature Communications*, vol. 14, 10 2023.
- [14] Z. Li, P. Wang, K. Liu, and Z. Tian, "MimoLoc: Indoor Localization With Assistance of Microwave Reflection of Downlink Signal in Sub-6G MIMO Networks," *IEEE Transactions on Microwave Theory and Techniques*, vol. 72, no. 4, pp. 2655–2668, Apr. 2024.
- [15] Z. Li, P. Wang, Z. Tian, and K. Liu, "TriLoc: Toward Accurate Indoor Localization With Assistance of Microwave Reflections," *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 6, pp. 2734–2747, Jun. 2023.
- [16] G. Zhang, D. Zhang, Y. He, J. Chen, F. Zhou, and Y. Chen, "Multi-Person Passive WiFi Indoor Localization With Intelligent Reflecting Surface," *IEEE Transactions on Wireless Communications*, vol. 22, no. 10, pp. 6534–6546, Oct. 2023.
- [17] I. An, Y. Kwon, and S.-e. Yoon, "Diffraction- and Reflection-Aware Multiple Sound Source Localization," *IEEE Transactions on Robotics*, vol. 38, no. 3, pp. 1925–1944, Jun. 2022.
- [18] L. I. Birnie, T. D. Abhayapala, and P. N. Samarasinghe, "Reflection Assisted Sound Source Localization Through a Harmonic Domain MUSIC Framework," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 279–293, 2020.
- [19] L. Yang, J. Liu, and X. Tang, "Depth from water reflection," *IEEE Transactions on Image Processing*, vol. 24, no. 4, pp. 1235–1243, 2015.
- [20] D. Bonilla Licea, V. Walter, M. Ghogho, and M. Saska, "Optical communication-based identification for multi-uav systems: theory and practice," *Autonomous Robots*, vol. 49, no. 3, p. 24, Sep 2025.
- [21] T. Lakemann, D. B. Licea, V. Walter, T. Báča, and M. Saska, "Towards agile multi-robot systems in the real world: Fast onboard tracking of active blinking markers for relative localization," *Robotics and Autonomous Systems*, vol. 194, p. 105175, 2025.
- [22] M. Born *et al.*, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed. Cambridge University Press, 1999.
- [23] E. Hecht, *Optics*, 5th ed. Boston: Pearson Education, 2017.
- [24] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion," in *Fourth IEEE International Conference on Computer Vision Systems (ICVS'06)*, 2006-01, pp. 45–45.
- [25] D. Hert, T. Baca, P. Petracek *et al.*, "MRS Drone: A Modular Platform for Real-World Deployment of Aerial Multi-Robot Systems," *Journal of Intelligent & Robotic Systems*, vol. 108, no. 4, p. 64, Aug. 2023.
- [26] A. Barišić, F. Petric, and S. Bogdan, "Brain over brawn: Using a stereo camera to detect, track, and intercept a faster UAV by reconstructing the intruder's trajectory," *Field Robotics*, vol. 2, pp. 222–240, 2022.