Drones and Autonomous Vehicles

SCIEPublish

Article

Camera View Selection for Bearing-Only Geolocation of Stationary Targets Using Multi-UAV Swarms with GPS Bias

Sebastian Quispe Arias *, Laercio Lucchesi, Tatiana Reimer, Thiago Lamenza, Bruno Olivieri and Markus Endler

Department of Informatics, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rua Marquês de São Vicente, 225, Gávea, Rio de Janeiro 22451-900, RJ, Brazil; llucchesi@inf.puc-rio.br (L.L.); tbarata@inf.puc-rio.br (T.R.); tlamenza@inf.puc-rio.br (T.L.); bolivieri@inf.puc-rio.br (B.O.); endler@inf.puc-rio.br (M.E.)

* Corresponding author. E-mail: sarias@inf.puc-rio.br (S.Q.A.)

Received: 1 June 2026; Revised: 1 July 2026; Accepted: 25 August 2026; Available online: 11 September 2026

ABSTRACT: We address the problem of geolocating a static ground target from multiple low-cost Unmanned Aerial Vehicles (UAVs) equipped with monocular cameras and object detectors. Each drone generates large sequences of bearing-only measurements per flyover, but these are affected by per-drone GPS biases (systematic position errors specific to each UAV) that remain effectively constant over short trajectories. We propose a greedy selector that combines angular diversity and detection confidence to pre-filter measurements before Random Sample Consensus (RANSAC)-based robust fusion. Across 30 randomized simulation runs, the selector achieves mean accuracy comparable to full-pool RANSAC (2.34 m *vs.* 2.44 m) while reducing median processing time by 4.1× (393 ms *vs.* 1602 ms). Similar accuracy is observed against LO-RANSAC and PROSAC, with the selector remaining 4× to 35× faster. The selector also outperforms confidence-only and standard Fisher Information Matrix (FIM)-based subset baselines. A bias-aware FIM analysis via Schur-complement marginalization shows that the information contributed by repeated measurements from a single drone saturates under a shared position bias, explaining the single-drone concentration observed for these baselines. In the nominal setting, the angular-diversity proxy retains 99.7% of the bias-aware FIM objective and achieves localization error similar to direct FIM-based selection, while being 3.3× faster at the selection stage. Real flights with a low-cost quadrotor measure short-term GPS drift of 1.9–3.5 m over 10–63 min, supporting the assumed bias dynamics, and expose a failure mode in which persistent false positives capture the RANSAC consensus. Full-pool RANSAC shows slightly lower mean error at larger GPS-bias levels ($\sigma_{GPS} \geq 5$ m), indicating that pre-selection is not preferable in every operating regime.

Keywords: UAV geolocation; Bearing-only localization; Fisher Information Matrix; View selection; RANSAC; Drone swarms; Search and rescue (SAR)

1. Introduction

Search and rescue (SAR) operations increasingly deploy multiple drones (UAVs) equipped with cameras to locate missing persons in disaster-affected areas. Once a candidate target is detected, its geographic coordinates must be estimated with sufficient accuracy to guide ground teams or to request further aerial support. Converting a detection into target coordinates requires knowing both the drone's position and its orientation. Consumer-grade drones rely on GPS receivers with horizontal errors on the order of meters and on magnetometer-based compasses, which are prone to orientation errors [1], thereby limiting the achievable geolocation accuracy.

Despite these limitations, multi-drone deployments (*i.e.*, drone swarms) can compensate by aggregating many observations. Visual detection identifies potential targets in each camera frame using an object detector (e.g., YOLO [2]). Each detection can be converted into a line of sight from the drone toward the ground using the pinhole camera model and the drone's reported GPS position and heading [3]. Multiple lines of sight from different viewpoints can be intersected to triangulate the target position [4]. When multiple drones observe the same target during a flyover, large sequences of bearing measurements accumulate in seconds. Not all of these measurements are equally useful, and selecting the right subset to locate the target is the focus of this work.

Beyond per-detection uncertainty from low-confidence observations, measurements from the same drone share a common GPS position bias that persists throughout the flight. Empirical studies have shown that consumer GPS errors are strongly temporally autocorrelated, with the autocorrelation decaying over tens of minutes [5], and that magnetometer heading biases can reach up to several degrees due to magnetic disturbances from the drone's motors [6]. These per-drone biases, therefore, constrain how a useful measurement subset can be chosen: large sequences of observations from a single drone cannot be treated as independent samples.

This bias structure directly affects measurement selection. Strategies that prioritize detection confidence tend to concentrate measurements on the closest drone, inheriting its GPS bias. A more principled approach would be to use the Fisher Information Matrix (FIM), which quantifies how much each measurement contributes to estimation accuracy. However, under independent-noise assumptions, FIM-based selection also favors the drone with the strongest sensitivity, producing the same over-concentration when per-drone biases are present.

Each of these challenges has been studied individually. Prior work has analyzed FIM-optimal observation geometries and submodular sensor selection under independent noise assumptions [7–9], including FIM-based formation design for cooperative UAV perception [10], explored confidence-aware noise modeling in two-dimensional tracking [11–13], and demonstrated single- and multi-drone geolocation from visual detections [1,14,15]. A particularly relevant prior approach is Kaplan's greedy FIM-based bearing-only node selection [16], which assumes independent bearing errors across nodes and does not address per-drone correlated bias or detection confidence.

This paper addresses the interaction between multi-UAV view selection and per-drone correlated GPS bias. We make three contributions:

1. A computationally efficient greedy selector that scores measurements by angular diversity and detection confidence. Used as a pre-filter for RANSAC, it achieves mean accuracy comparable to full-pool RANSAC in the main configuration (2.34 m *vs.* 2.44 m), provides a 4.1× median processing time speedup (393 ms *vs.* 1602 ms), and improves over confidence-only and standard-FIM subset baselines. Unlike random subsets, the selection step is deterministic, avoiding the multi-meter accuracy spread that random draws exhibit across runs.
2. A bias-aware Fisher Information analysis via Schur-complement marginalization, showing that each drone's information contribution saturates after a small number of measurements per drone. This

explains the empirical single-drone concentration observed in standard FIM-based and confidence-only selection under per-drone bias.

3. A diagnostic study with per-drone selection statistics that delineates the operating regimes of greedy view selection, showing where it is competitive and where full-pool RANSAC may be preferred. This study identifies larger GPS bias as the key limitation, and characterizes how lateral target offsets reduce viewing diversity, a geometric effect whose impact on accuracy was not statistically significant in a 300-seed re-run.

In addition, a real-flight evaluation (Section 5) tests the central bias assumption on hardware and documents a multi-object failure mode that limits unattended use.

2. Related Work

Estimating a target's position from angular measurements alone, without range information, is known as bearing-only localization. Classical approaches use weighted and total least squares formulations [4,17]. The role of observation geometry in estimation quality has been characterized through the Fisher Information Matrix [7], and the submodularity of log det(F) enables greedy sensor selection with a $(1 - 1/e)$ approximation guarantee [8,18]. More broadly, sensor selection has been addressed through convex optimization [19] and informative placement with submodular objectives [20]. Closer to UAV applications, FIM determinant maximization has been applied to design information-optimal formations for multimodal UAV cooperative perception [10]. Particularly relevant to bearing-only selection, Kaplan [16] proposed greedy FIM-based node selection for bearing-only localization, assuming independent bearing errors across nodes and without modeling per-detection confidence. Kirchner et al. [9] extended submodular FIM selection to heterogeneous sensor types, with measurement-noise models tied to sensor type rather than derived from per-measurement detection confidence. Across these works, measurement errors are treated as independent across nodes or platforms; per-drone correlated bias is not modeled.

A separate line of research has explored how detection quality itself can inform estimation. Detection confidence from modern object detectors has been used to modulate observation noise in state estimation. Examples include propagating bounding-box covariance into Kalman filters [11], deriving per-detection noise from YOLO confidence [12], and using confidence for track management [13]. However, these approaches operate in two-dimensional image-plane tracking with associated trajectories; they do not address the selection of measurements from a pool of bearing observations for three-dimensional geolocation.

UAV-based geolocation has progressed from single-drone systems achieving 1–10 m accuracy with YOLOv8 [1] and raycasting-based SAR geolocation [14], to multi-drone cooperative approaches. These include end-to-end visual geolocation from multiple micro-UAVs [21], FIM-guided bearing-only geolocation of moving targets from micro-UAVs [22], and trajectory optimization for bearing-only localization using the posterior Cramér-Rao bound in multi-drone settings [15]. At the application level, consumer-grade object geolocation pipelines have been demonstrated [3], and a recent review covers sensing technologies for UAV swarms in SAR missions [23]. These works focus on detection accuracy, trajectory optimization, or pipeline-level demonstrations; measurement selection that accounts for per-drone correlated bias is not addressed.

Once measurements are selected, a robust estimator is needed to handle the remaining outliers. RANSAC [24] remains the dominant approach for robust estimation in the presence of outliers, with extensions such as PROSAC [25] for confidence-ranked sampling, LO-RANSAC [26] for local optimization of promising hypotheses, and USAC [27] as a unified framework. Our approach is complementary: we pre-filter the measurements to improve both quality and source diversity before RANSAC.

3. System Model and Method

The method comprises four components: a measurement model with per-drone biases, a Fisher Information analysis that motivates cross-drone diversification, a greedy view selector, and an RANSAC-based fusion stage. We consider multiple drones flying in formation over a search area at a fixed altitude. A static target is located at an unknown position $\mathbf{q}^* = (x^*, y^*, 0)$ on the ground plane. Each drone carries a monocular camera at a fixed off-nadir pitch of −75° and runs an object detector. Figure 1 illustrates the scenario.

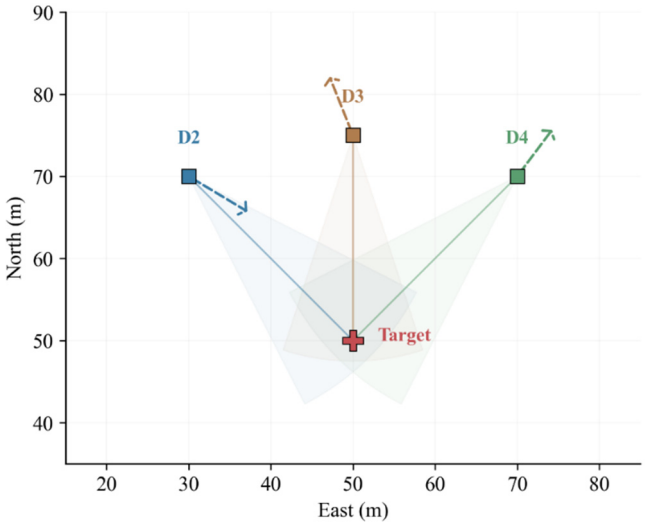

Figure 1. Multi-UAV bearing-only geolocation scenario. Three drones in V formation observe a ground target. Each drone has a fixed GPS bias (dashed arrows) shared by all its measurements.

Each detection produces a bearing-only measurement $m_i = (\hat{\mathbf{p}}_i, \hat{\mathbf{d}}_i)$: a reported drone position and a ray direction. Each detection center is back-projected with the camera intrinsic matrix and rotated into the world frame using the camera attitude to obtain the bearing vector [3,28]. Let $j(i)$ denote the drone that produced measurement i. GPS and heading errors are modeled as per-drone biases that remain effectively constant throughout a flyover:

$$\begin{aligned} \hat{\mathbf{p}}_i &= \mathbf{p}_i + \boldsymbol{\delta}^{\mathrm{GPS}}_{j(i)}, \quad \boldsymbol{\delta}^{\mathrm{GPS}}_j \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\mathrm{GPS}}\mathbf{I}_2) \\ \hat{\psi}_i &= \psi_i + \delta^{\psi}_{j(i)}, \quad \delta^{\psi}_j \sim \mathcal{N}(0, \sigma^2_{\psi}) \end{aligned} \tag{1}$$

where $\mathbf{p}_i$ and ψ_i are the true position and heading at measurement i, and $\boldsymbol{\delta}^{\mathrm{GPS}}_j$ and δ^{ψ}_j are biases drawn once per drone and shared by all its measurements. This is motivated by GPS autocorrelation that persists for tens of minutes [5] and magnetometer biases due to motor interference reaching several degrees [6]. For a 30-s flyover, these errors are effectively constant.

Pixel noise is independent per detection, with a standard deviation $\sigma_{\mathrm{px},i} = C/c_i$, where C is a calibration constant and c_i the detection confidence. Because detector class confidence is not necessarily equivalent to bounding-box localization quality [29], this model is treated as a calibrated proxy rather than a precise noise estimate. The confidence-to-noise relationship was empirically calibrated on 3014 real detections from the VisDrone2019 aerial dataset [30], where ground-truth bounding boxes are available. For each detection, pixel localization error was defined as the Euclidean distance between the predicted and ground-truth box centers, and it was compared with detector confidence using Spearman's rank correlation. The resulting negative correlation (Spearman $\rho_s = -0.287$, $p \ll 0.001$) indicates that higher-confidence detections tend to have lower localization error, supporting the use of confidence as a noise-weighting signal.

The sensitivity of the projected pixel coordinates to changes in the target position is captured by the 2×2 Jacobian matrix $\mathbf{J}_i$, computed numerically via small perturbations through the pinhole model. Under

independent pixel noise, each measurement contributes $\mathbf{F}_i = w_i \mathbf{J}_i^T \mathbf{J}_i$ to the Fisher Information Matrix, where $w_i = (c_i/C)^2$. The standard FIM sums these contributions assuming independence: $\mathbf{F}_{\text{std}} = \sum_i \mathbf{F}_i$. Greedy maximization of $\det(\mathbf{F}_{\text{std}})$ concentrates all selections on the drone with the largest Jacobians, inheriting its GPS bias.

To account for per-drone GPS bias, we marginalize the translational GPS bias and treat it as a nuisance parameter. This marginalization is restricted to translation: GPS bias shifts the reported camera position, so its local effect can be represented as a translational nuisance term in the projection model. Yaw bias, by contrast, rotates the bearing direction through the camera-attitude model, making its effect geometry-dependent and nonlinear. Including yaw in the same FIM derivation would require an additional attitude linearization or a numerical FIM, so we evaluate yaw sensitivity empirically instead. The pinhole projection for measurement i depends on $\mathbf{q} - \mathbf{p}_i$; therefore, shifting the drone trajectory by $+\boldsymbol{\delta}_{j(i)}^{\text{GPS}}$ is locally equivalent to shifting the target by $-\boldsymbol{\delta}_{j(i)}^{\text{GPS}}$, yielding $\mathbf{J}_{\text{bias}} = -\mathbf{J}_{\text{target}}$ for translational GPS bias. For the subset $\mathcal{S}_j$ of measurements produced by drone j, let $\mathbf{S}_j = \sum_{i \in \mathcal{S}_j} w_i \mathbf{J}_i^T \mathbf{J}_i$ be the accumulated FIM. The joint FIM for target and bias is:

$$\mathbf{F}_{\text{joint},j} = \begin{bmatrix} \mathbf{S}_j & -\mathbf{S}_j \\ -\mathbf{S}_j & \mathbf{S}_j + \sigma_{\text{GPS}}^{-2}\mathbf{I} \end{bmatrix} \tag{2}$$

Marginalizing the bias via the Schur complement yields:

$$\mathbf{F}_{\text{eff},j} = \mathbf{S}_j - \mathbf{S}_j\left(\mathbf{S}_j + \sigma_{\text{GPS}}^{-2}\mathbf{I}\right)^{-1}\mathbf{S}_j \tag{3}$$

As $\mathbf{S}_j$ grows, $\mathbf{F}_{\text{eff},j} \to \sigma_{\text{GPS}}^{-2}\mathbf{I}$. With $\sigma_{\text{GPS}} = 3$ m, each drone contributes at most $(1/9)\mathbf{I}$ regardless of measurement count. Figure 2 illustrates this: the standard FIM determinant grows from 508 to over 5×10^6 as measurements increase from 1 to 100, while the bias-aware FIM determinant saturates at 0.0123 after a single measurement. Consequently, same-drone measurements provide diminishing information once the bias ceiling is reached, whereas measurements from another drone add non-redundant information. The Cramér-Rao Lower Bound, obtained by inverting the total bias-aware FIM, provides a reference scale under the assumed unbiased estimation model:

$$\text{CRLB}_{\text{RMSE}} = \sqrt{\text{tr}\left(\mathbf{F}_{\text{total}}^{-1}\right)} \tag{4}$$

In this work, this bound is used as a diagnostic scale for redundancy under the bias-aware model; its numerical value and relationship to the proposed selector's empirical performance are examined in Section 4.

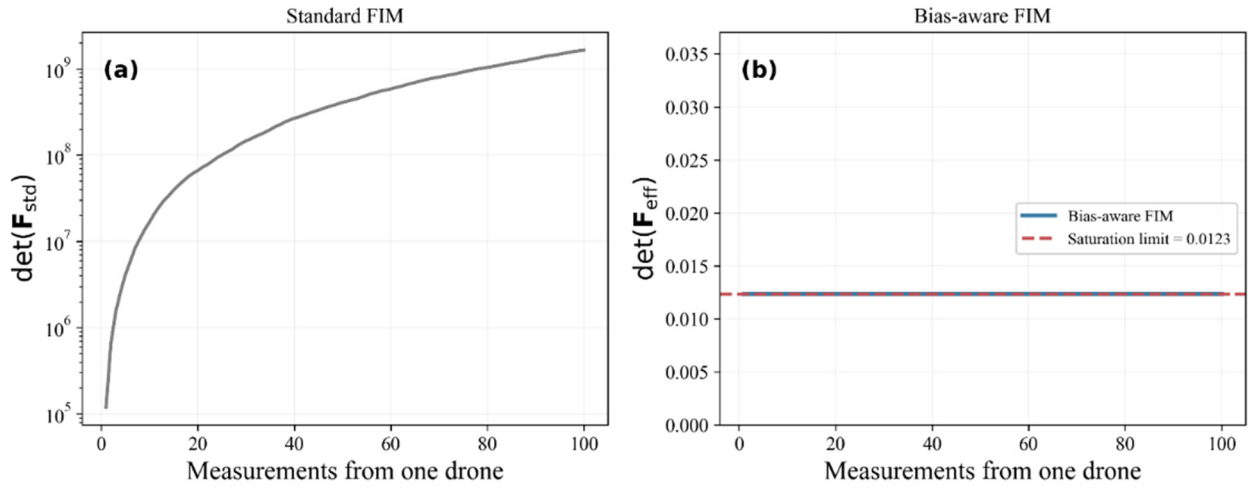

Figure 2. Information saturation under per-drone GPS bias: (**a**) The standard FIM grows without saturating as more measurements are added; (**b**) The bias-aware FIM saturates after a few measurements, motivating cross-drone diversification.

$$s_i = \alpha \cdot \frac{\delta_i}{90^\circ} + (1-\alpha) \cdot c_i \tag{5}$$

where $\delta_i = \min_{s \in S} \arccos\left(\left|\hat{\mathbf{d}}_i \cdot \hat{\mathbf{d}}_s\right|\right)$ is the angular diversity relative to the current selection, c_i is the detection confidence, and $\alpha = 0.2$. The selection is initialized with the highest-confidence measurement and proceeds greedily, rejecting candidates below a minimum angular threshold $\delta_{\min} = 10^\circ$ (Algorithm 1). Although the default budget is $K = 30$, the minimum-angle filter causes the selector to self-limit to approximately 12 effective measurements, as same-drone candidates produce nearly parallel rays under near-nadir geometry and are progressively rejected. This self-limiting behavior avoids the redundant same-drone measurements that the FIM saturation property identifies as uninformative. The selector thus returns up to K measurements rather than exactly K.

The angular threshold $\delta_{\min} = 10^\circ$ acts as a default redundancy-control parameter. In the default setup, temporally adjacent bearings from the same drone are nearly parallel (median consecutive separation $\approx 0.5^\circ$), so the threshold suppresses repeated same-drone measurements while retaining sufficiently separated views for RANSAC. It is not intended to resolve all geometric degeneracies: under lateral target offsets, views from different drones become more same-sided, which motivates the lateral-offset analysis in Section 4.3.

After the angular filter enforces a baseline level of diversity, the confidence-geometry weight α ranks the remaining candidates. The default $\alpha = 0.2$ lies in a flat, statistically stable region of the response curve, while the main diversification mechanism is the angular constraint rather than a sharply tuned confidence-geometry weight. At each iteration, the algorithm first removes candidates that are too close in bearing angle to the current selection and then ranks the remaining candidates using the combined confidence-diversity score.

Algorithm 1. Greedy view selection.

1: Input: Measurements M, confidences $\{c_i\}$, maximum budget K, balance α, angular threshold δ_{min}
2: Initialize: S ← {argmax$_i$ c_i} // highest-confidence measurement first
3: while |S| < K do
4: C ← ∅ // feasible candidates
5: for each i ∉ S do
6: δ_i ← min$_s$ ∈ S arccos(|$\hat{d}_i \cdot \hat{d}_s$|) // |·|: antiparallel ≈ redundant, δ_i ∈ [0°, 90°]
7: if $\delta_i \geq \delta_{min}$ then
8: s_i ← α * (δ_i/90°) + (1 − α) * c_i // δ_i/90° normalizes separation to [0, 1]
9: C ← C ∪ {i}
10: end if
11: end for
12: if C = ∅ then break end if // self-limiting stop (see text)
13: S ← S ∪ {argmax$_i$∈C s_i } // add best candidate
14: end while
15: return S
Output: Selected subset S of bearing-only measurements

The absolute value in the separation (Algorithm 1, line 6) treats near-antiparallel rays as redundant, bounding δ_i by 90° and motivating its normalization in the score. When no candidate clears $\delta_{\min}$, selection stops early (line 12), which caps the effective subset at ~12 measurements independently of K.

The selected measurements are passed to RANSAC for robust fusion. In each of 100 iterations, two rays are sampled uniformly at random and used to form a candidate target estimate from their closest point of approach. Each remaining measurement is then scored by its ray-to-point residual to the candidate estimate, and measurements with a residual below τ = 5 m are counted as inliers. The final estimate is

obtained by ground-constrained gradient descent over the selected inlier set, minimizing a uniformly weighted sum of squared ray-to-point distances with the target constrained to the ground plane ($z = 0$). Thus, confidence scores are used for measurement pre-selection, but not for reweighting the final RANSAC inlier refinement. Since the target is assumed to lie on the ground plane, the refinement estimates only its horizontal coordinates. Figure 3 summarizes the pipeline.

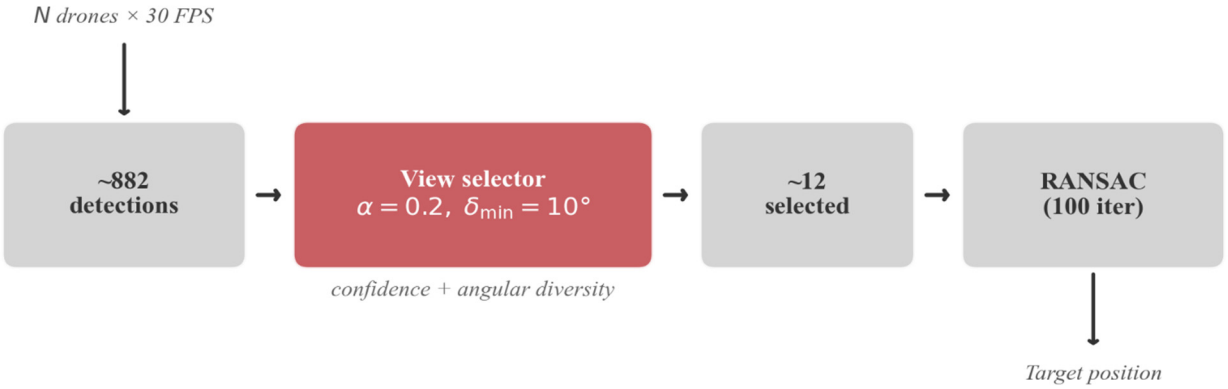

Figure 3. Two-stage pipeline. The selector reduces the measurement pool to a compact subset, then RANSAC performs robust fusion with ground-constrained re-estimation.

4. Experimental Evaluation

Experiments are conducted in GrADyS-Sim NextGen [31], a discrete-event multi-agent simulator. Table 1 summarizes the default parameters. The evaluation covers the main comparison under the default configuration, sensitivity to selector and noise parameters, practical perturbation tests, and robustness across formation and target-placement changes, including a diagnostic analysis of the dominant failure mode. Twelve empirical methods are compared: seven robust RANSAC-based methods (the proposed selector and six baselines), two FIM-based selectors for the theoretical context, and three non-robust baselines, with a bias-aware CRLB reported as reference. Non-robust baselines use all selected measurements directly, without RANSAC, and apply equal-weight gradient-descent fusion constrained to the ground plane.

4.1. Main Comparison

In the main configuration, the proposed selector achieves a mean error of 2.34 m, comparable to the 2.44 m of full-pool RANSAC (Wilcoxon $p = 0.38$; paired W/L = 11/19), with similar median error (2.23 m *vs.* 2.14 m). An equivalence test (TOST) further shows that the paired difference lies within ±0.50 m ($p = 0.007$), a small margin relative to the meter-level errors in this setting. Table 2 reports the main comparison and diagnostic baselines. Compared to other RANSAC-based subset baselines, the selector is significantly more accurate than confidence-only selection (4.20 m, $p = 0.001$). Random subset selection reaches 2.85 m with a lower runtime (30 ms); its difference from the proposed selector ($p = 0.043$) does not survive multiple-comparison (Holm) correction. The practical drawback of random selection is nondeterminism, which the deterministic selection step avoids: individual draws vary by a median of 4.3 m, about 69% of draws are worse than the proposed selector, and the tail is heavier (maximum error 7.42 m *vs.* 5.29 m in Table 2). Against stronger full-pool robust estimators, the selector remains comparable to PROSAC (2.52 m, $p = 0.58$) and LO-RANSAC (2.43 m, $p = 0.39$), while running 35× faster than LO-RANSAC. In summary, the proposed selector achieves accuracy comparable to full-pool RANSAC while running 4.1× faster (1602 ms → 393 ms median). The greedy selection itself accounts for 378 ms; the subsequent RANSAC step on the reduced subset adds only 15 ms, compared to 1602 ms when RANSAC operates on

the full pool of approximately 882 measurements. The reduction follows from the algorithm rather than from the implementation: each RANSAC iteration scores every measurement in the pool, so the fusion cost grows with the pool size, and all methods share the same RANSAC implementation.

Among FIM-based selectors, the bias-aware variant at $K = 20$ (chosen to match the saturation regime identified by the FIM analysis) achieves 2.44 m, comparable to the proposed selector ($p = 0.47$) and full-pool RANSAC; all three approaches lie near the 2.45 m diagnostic scale estimated by the bias-aware FIM model. In contrast, the standard FIM produces 4.18 m, significantly worse than the proposed selector ($p = 0.001$, 23W/7L) and substantially worse than the bias-aware variant, supporting the interpretation that ignoring per-drone bias can lead to degraded accuracy through single-drone concentration, a pattern also observed for the confidence-only selector (Table 3). The non-robust uniform-averaging baseline also reaches a similar mean error (2.47 m), consistent with the FIM saturation analysis: with many cross-drone measurements, even simple averaging approaches the diagnostic scale in the main setting. However, it requires fusing the full measurement pool and runs substantially slower than the proposed selector (1228 ms *vs.* 393 ms median).

To contextualize these results, the bias-aware FIM analysis gives a diagnostic scale of 2.45 m for the default configuration. The value computed from all approximately 882 measurements (2.450 m) is nearly identical to that obtained from the approximately 12 selected by the proposed selector (2.452 m), a difference of 0.003 m (0.1%). This supports the redundancy interpretation of the FIM analysis: additional same-drone measurements contribute little independent information under the per-drone bias model. This scale is used as a diagnostic reference for redundancy under the bias-aware model, rather than as a strict lower bound on the observed mean error. The empirical 2.34 m result lies close to this diagnostic scale, consistent with the selector operating near the redundancy limit of the bias-aware model.

Table 1. Default experimental configuration.

Parameter	Value	Description
Scenario		
N_d	3	UAVs in V formation, 20 m inter-UAV separation
Altitude	50 m	Above ground level
Speed	5 m/s	Horizontal flight speed
Duration	30 s	Simulated flyover duration
Target position	(50,50,0) m	Default ground-truth target position
Ground constraint	$z = 0$	Ground target assumption
Sensor and detection		
Camera	SIYI A8 mini	1920×1080, $f \approx 1130$ px
Camera pitch	-75°	Fixed off-nadir gimbal angle
Detector	30 FPS, $c_i \geq 0.3$	Detection rate and confidence threshold
Noise model		
σ_{GPS}	3 m	Per-drone GPS bias standard deviation
σ_{ψ}	5°	Per-drone yaw bias standard deviation
Pixel noise	$\sigma_{\mathrm{px}} = 5.0/c_i$	Confidence-dependent detection noise
Selection and fusion		
K	30	Default view budget; selector returns up to K views
α	0.2	Score weight: 0.2 geometry, 0.8 confidence
$\delta_{\min}$	10°	Minimum angular separation between selected views
RANSAC	100 iter., $\tau = 5$ m	Robust fusion iterations and inlier threshold
Evaluation		
Seeds	0–29	30 Monte Carlo runs per condition

Italics indicate group headings, not parameters.

Table 2. Main comparison and diagnostic baselines over 30 seeds. Errors are reported in meters and runtimes in milliseconds.

Method	Mean	Std	Med	P90	Max	ms	Speedup
Robust methods							
Proposed selector	**2.34**	1.38	2.23	4.26	**5.29**	393	4.1×
Full-pool RANSAC	2.44	1.58	2.14	4.44	6.36	1602	1.0×
LO-RANSAC (full-pool)	2.43	**1.31**	2.14	**4.07**	5.57	13,807	0.1×
PROSAC (full-pool)	2.52	1.51	2.48	4.31	6.08	1560	1.0×
Top-K geometric subset	2.59	1.75	**2.04**	5.06	6.87	2425	0.7×
Random subset	2.85	1.70	2.17	4.91	7.42	30	52.8×
Confidence-only	4.20	2.73	3.70	8.11	11.41	31	52.6×
FIM-based selectors							
FIM bias-aware ($K = 20$)	2.44	1.59	2.20	4.73	5.78	898	1.8×
FIM standard ($K = 30$)	4.18	2.66	3.65	7.69	11.17	117	13.7×
Non-robust baselines							
Confidence-only (no RANSAC)	4.19	2.72	3.70	8.11	11.41	8	210.0×
Single closest drone	4.24	2.78	3.78	8.46	11.58	194	8.3×
Uniform averaging (all measurements)	2.47	1.54	2.31	4.99	5.81	1228	1.3×
Diagnostic reference							
Bias-aware CRLB	2.45	—	—	—	—	—	—

Italics indicate group headings. Bold indicates the lowest value in each error column. Em dashes indicate quantities that do not apply to the CRLB, which is a bound rather than an estimator.

Table 3. Average drone distribution of selected measurements (30 seeds, centered target). D2, D3, D4 denote the three drones in the V formation.

Method	D2	D3	D4	Total
Proposed selector	3.7	4.4	3.7	~12
FIM bias-aware ($K = 20$)	7.1	5.9	7.0	20
FIM standard	0.0	30.0	0.0	30
Confidence-only	0.0	30.0	0.0	30

The default-regime results in Tables 2 and 3 show that the proposed angular-diversity selector and the bias-aware FIM selector both avoid the single-UAV concentration observed for standard FIM and confidence-only selection. To test whether this relationship holds beyond the default configuration, we compared the proposed selector with explicit bias-aware FIM selectors across five operating regimes: the centered target with σ_{GPS} = 3 m, increased GPS-bias levels of 5 m and 8 m, and left/right lateral target offsets (Table 4). In the nominal regime, the proposed selector remains comparable to the FIM-bias $K = 20$ and $K = 30$, while the standard FIM produces substantially larger error. This is consistent with angular diversity capturing the main diversification effect predicted by the bias-aware FIM analysis. At the objective level, the proxy is also near-optimal: subsets chosen by angular diversity attain a median 99.7% of the bias-aware FIM objective reached by direct greedy maximization in the nominal setting, with no statistically significant accuracy difference (2.34 m *vs.* 2.39 m at $K = 30$, paired Wilcoxon $p = 0.23$) and 3.3× lower selection cost. However, the proxy is not equivalent to explicit FIM optimization. At higher GPS bias and for the left-offset target, the bias-aware FIM selector obtains lower mean error, although the paired differences are not statistically significant across the GPS-noise cases (p = 0.382–0.903) and show only near-threshold evidence for the left-offset case ($p = 0.061$). The proposed selector is therefore best interpreted as a computationally simpler pre-selection method with lower runtime, while explicit bias-aware FIM selection may be preferable when reliable noise parameters are available, or the operating regime is more challenging. In deployment this condition is hard to meet: bias-aware selection needs the per-drone

noise scales as inputs, and the real flights in Section 5 show the effective GPS bias changing between 1.9 and 3.5 m depending on the time window, while the angular proxy needs no noise parameters.

Table 4. Angular proxy versus bias-aware FIM across operating regimes. Mean errors are reported in meters over 30 seeds; runtime ratio is computed relative to the proposed selector.

Condition	Proposed	FIM-Bias K = 20	FIM-Bias K = 30	Standard FIM	Full-Pool	*p vs.* FIM-Bias K = 20	Runtime Ratio
Center, GPS = 3 m	2.34	2.44	2.39	4.18	2.44	0.465	2.4×
Center, GPS = 5 m	4.77	4.32	4.14	6.97	4.70	0.382	2.4×
Center, GPS = 8 m	9.06	7.73	7.44	11.17	8.68	0.903	2.3×
Left target	3.27	2.81	2.84	4.09	2.92	0.061	2.6×
Right target	2.36	2.37	2.17	3.90	2.34	0.440	2.5×

4.2. Sensitivity Analysis

Table 5 consolidates the parameter sensitivity analysis. The tested selector hyperparameters fall in stable regions of their response curves. The geometry-confidence balance α = 0.2 lies within a flat, statistically stable region of the parameter curve: values in [0.1, 0.6] produce comparable mean errors with no statistically significant difference (paired Wilcoxon $p = 0.43$ for $\alpha = 0.2$ *vs.* $\alpha = 0.6$). We retain $\alpha = 0.2$ as the default throughout this work. The selector requires $K \geq 20$ to reach the main-configuration mean accuracy comparable to full-pool RANSAC; smaller budgets ($K = 10$) lose 0.6 m on average because the algorithm terminates before reaching its self-limiting ~12 effective measurements. The remaining default thresholds were also checked empirically. For the angular threshold, $\delta_{\min} = 10°$ achieved the lowest mean error among the tested values {5°, 10°, 15°, 20°}, while $\delta_{\min}$= 5° selected more redundant views and was substantially slower. For the RANSAC inlier threshold, $\tau = 5$ m achieved the lowest mean error for the proposed selector among the tested values {3, 5, 7, 10 m}; $\tau = 3$ m was too strict and increased the mean error to 3.26 m. Two quantities were held fixed throughout: the RANSAC iteration count (100) and the confidence-to-noise constant (C = 5 in σ_{px} = C/confidence); the latter mapping is supported qualitatively by our real-flight data (Section 5), although the exact 1/confidence form is an approximation.

Sensitivity to simulation and noise parameters shows where the trade-off is stable and where it degrades. Under default noise ($\sigma_{GPS} = 3$ m), the selector and full-pool RANSAC show no statistically significant difference (paired Wilcoxon $p = 0.38$). As shown in Figure 4, full-pool RANSAC achieves slightly lower mean error than the selector at $\sigma_{GPS} = 5$ m (4.70 *vs.* 4.77 m) and $\sigma_{GPS} = 8$ m (8.68 *vs.* 9.06 m). Neither paired difference is statistically significant (Wilcoxon $p = 0.58$ and $p = 0.92$). The selector's restricted measurement subset likely cannot average correlated bias errors as effectively as the full pool when the per-drone GPS bias scale dominates. The number of drones reveals a distinct effect: with a single drone, there is no inter-drone diversity to exploit, and all methods perform similarly (the proposed selector and full-pool RANSAC both around 4.1 m). The selector's distinct behavior emerges only in the multi-drone regime ($N_d \geq 3$). Because yaw bias is not marginalized in the closed-form FIM, we evaluate its effect empirically. Yaw noise has a more modest impact: at low yaw (0–2°), full-pool RANSAC has slightly lower mean error; at higher yaw (10–15°), the selector is competitive or slightly better. No yaw level produces a statistically significant difference (paired Wilcoxon $p \geq 0.06$ across all five tested values), so the selector remains competitive and does not exhibit collapse despite the FIM model omitting yaw bias. Finally, the confidence-to-noise rule affects absolute error but preserves the comparison with full-pool RANSAC. Under the VisDrone-derived rule ($C = 1.13$), the selector's mean error is 2.56 m versus 2.49 m for full-pool RANSAC (paired Wilcoxon $p = 0.50$); the comparison thus remains non-significant despite the sign change relative to the baseline rule ($C = 5.0$, where the selector reaches 2.34 m). The confidence-only baseline remains substantially worse than both robust methods under both rules.

We also tested whether confidence should be used in the final refinement after RANSAC. In this variant, RANSAC sampling and inlier selection are unchanged; only the final ground-plane refinement is weighted by detection confidence. This did not improve the proposed selector. Mean error remained unchanged in the default setting (2.34 m *vs.* 2.34 m) and slightly increased at $\sigma_{\mathrm{GPS}} = 5$ m (4.77 m *vs.* 4.81 m), $\sigma_{\mathrm{GPS}} = 8$ m (9.06 m *vs.* 9.07 m), and for left/right target offsets (3.27 m *vs.* 3.35 m; 2.36 m *vs.* 2.40 m). None of these paired differences is statistically significant ($p \geq 0.253$). In a control case with no GPS or yaw bias, where pixel noise is the main error source, confidence weighting reduced full-pool RANSAC's mean error from 0.012 m to 0.011 m ($p = 0.002$) but left the proposed selector unchanged ($p = 0.46$). We therefore retain uniform refinement as the default, since in the evaluated regimes the remaining error after RANSAC is driven more by shared per-drone pose bias than by per-detection confidence differences. We revisited this question under adversarial contamination and on real flight data (Section 5): with injected false positives carrying realistically reduced confidence (scaled by 0.74, the empirical false-positive/target confidence ratio measured in our flights), the weighted refit remains statistically neutral at contamination levels up to 30% ($p \geq 0.37$ on paired runs with identical RANSAC draws), and a confidence-weighted vote performs slightly worse. On real data, weighting shifts a captured estimate toward the true target by up to 0.8 m but never reverses the capture. Confidence is thus most effective at the time of selection.

Table 5. Sensitivity analysis of the proposed selector. Mean errors are reported in meters; each row varies one parameter while all others remain fixed. Dashes indicate settings not evaluated for the corresponding baseline.

Parameter	Value	Proposed	Full-Pool	Conf-Only
α	0.0	2.64	**2.44**	—
α	0.1	2.45	**2.44**	—
α	0.2	**2.34**	2.44	—
α	0.4	2.48	**2.44**	—
α	0.6	**2.18**	2.44	—
K	10	2.94	**2.44**	4.19
K	20	**2.34**	2.44	4.19
K	30	**2.34**	2.44	4.20
K	50	**2.34**	2.44	4.19
σ_{GPS}	3 m	**2.34**	2.44	4.20
σ_{GPS}	5 m	4.77	**4.70**	6.97
σ_{GPS}	8 m	9.06	**8.68**	11.18
N_d (drones)	1	**4.12**	4.17	4.13
N_d (drones)	3	**2.34**	2.44	4.20
N_d (drones)	5	2.11	**2.01**	3.95
σ_ψ	0°	2.47	**2.32**	4.24
σ_ψ	2°	2.57	**2.45**	4.19
σ_ψ	5°	**2.34**	2.44	4.20
σ_ψ	10°	**3.10**	3.11	4.49
σ_ψ	15°	**3.70**	4.01	5.00
Confidence	Baseline($C = 5.0$)	**2.34**	2.44	4.20
Confidence	VisDrone($C = 1.13$)	2.56	**2.49**	4.20

Bold indicates the lowest mean error in each row. Em dashes indicate that confidence-only selection does not depend on α.

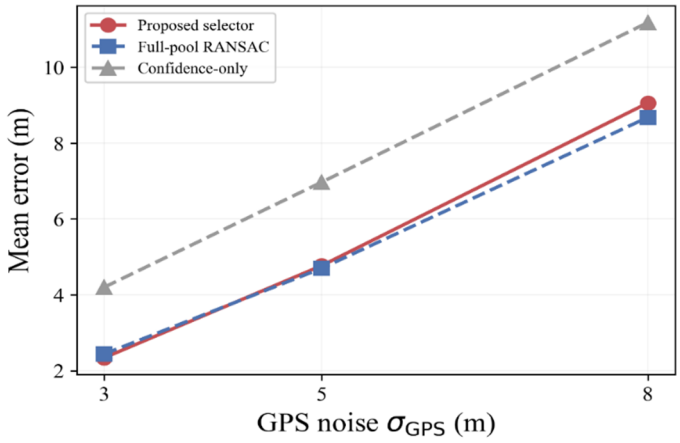

Figure 4. Mean error as a function of GPS noise for the proposed selector, full-pool RANSAC, and confidence-only selection.

In addition to the nominal parameter sweeps in Table 5, we tested practical perturbations that can occur in UAV deployment. These perturbations are applied as mismatches between the measurements generated by the simulator and the parameters assumed by the estimation pipeline. They include camera intrinsic miscalibration, gimbal pitch bias, pose-timestamp mismatch, detection dropout, terrain-height mismatch, per-drone asynchronous pose offsets, and increased detection noise. These tests remain simulation-based and are intended to quantify sensitivity to common deployment errors under controlled conditions.

Table 6 shows that the proposed selector and full-pool RANSAC degrade similarly across the tested single-perturbation conditions. The paired differences are not statistically significant across these conditions (Wilcoxon $p \in [0.152, 0.808]$), indicating no evidence that view selection is disproportionately fragile under the tested perturbations. Confidence-only selection remains consistently worse, with mean errors between 4.12 m and 6.94 m.

Terrain-height mismatch is the strongest tested perturbation. Because the estimator is constrained to the assumed ground plane, terrain mismatch affects not only the vertical component of the 3D error but also the horizontal geolocation estimate. For the proposed selector, a 5 m height mismatch increases the total 3D error to 6.09 m and the horizontal XY error from 2.34 m to 3.17 m. This confirms that the flat-ground assumption is a substantive limitation rather than a purely vertical-error issue.

Table 6. Practical perturbation tests over 30 seeds. Mean errors are reported in meters, and *p* denotes the paired Wilcoxon test between the proposed selector and full-pool RANSAC.

Condition	Perturbation	Proposed	Full-Pool	Conf-Only	*p*
Baseline	none	2.34	2.44	4.20	0.382
Intrinsics mild	f + 5%, pp + 10 px	2.53	2.42	4.12	0.570
Intrinsics strong	f + 10%, pp + 20 px	2.50	2.49	4.17	0.641
Gimbal mild	pitch + 1°	2.64	2.67	4.27	0.477
Gimbal strong	pitch + 3°	2.95	2.83	4.77	0.184
Time sync mild	100 ms	2.48	2.56	4.28	0.490
Time sync strong	300 ms	2.98	3.03	4.58	0.598
Dropout mild	30% removed	2.52	2.45	4.21	0.360
Dropout strong	50% removed	2.57	2.44	4.19	0.404
Terrain mild	target z = 2 m	3.40	3.63	4.86	0.229
Terrain strong	target z = 5 m	6.09	6.12	6.94	0.171
Async mild	±100 ms/drone	2.50	2.37	4.25	0.152
Async strong	±300 ms/drone	2.30	2.33	4.41	0.808
Detection noise mild	extra C = 10	2.45	2.49	4.21	0.761
Detection noise strong	extra C = 20	2.49	2.54	4.22	0.452

4.3. Robustness and Lateral-Offset Diagnosis

The robustness analysis, summarized in Table 7, evaluates whether the main speed/accuracy trade-off persists across formation geometries and target placements. Across all three formations, we do not observe statistically significant paired differences between the selector and full-pool RANSAC (paired Wilcoxon $p \geq 0.38$ for line, V, and triangle). Line and V formations slightly favor the selector in mean error (2.31 m *vs.* 2.49 m for line; 2.34 m *vs.* 2.44 m for V), while triangle slightly favors full-pool RANSAC (2.54 m *vs.* 2.47 m); none of these differences are significant. The W/L counts also vary modestly across formations (14/16, 11/19, 15/15). Target placement shows a similar pattern for along-track displacement: at near (2.48 m *vs.* 2.44 m, $p = 0.26$) and far (2.33 m *vs.* 2.74 m, $p = 0.36$) positions, neither method shows a significant paired advantage. Lateral displacement produces the largest observed gap: for the left-offset target, the proposed selector is 0.35 m worse than full-pool RANSAC (3.27 m *vs.* 2.92 m, paired Wilcoxon $p = 0.033$, 10W/20L). This is the only formation and target-position condition in Table 7 where the two methods differ at uncorrected conventional significance. The mirror right-offset condition produces nearly equal mean errors (2.36 m *vs.* 2.34 m, $p = 0.44$, 19W/11L). Because the left and right configurations are geometrically symmetric, the observed left/right discrepancy is best interpreted as finite-sample sensitivity to the realized per-drone GPS biases rather than as a deterministic geometric asymmetry. This motivates the diagnostic analysis below.

To identify the mechanism underlying the observed left-offset imbalance, we conducted a diagnostic study in which noise sources were individually removed (Table 8). Figure 5 illustrates the geometric change: when the target is centered, the three drones observe from well-separated bearings and the selector distributes views across all drones (D2: 3.7, D3: 4.4, D4: 3.7 measurements on average); when the target is laterally offset to the left, the geometry provides less diverse viewing directions and the far drone becomes underrepresented (D2: 4.3, D3: 4.0, D4: 2.7), reducing cross-drone balance and limiting the averaging of per-drone GPS bias realizations. Removing GPS bias eliminates and reverses the baseline disadvantage: the selector achieves a small, nominally significant improvement over full-pool RANSAC (1.35 m *vs.* 1.38 m, paired Wilcoxon $p = 0.043$, 21W/9L). Removing yaw bias alone preserves most of the gap (2.68 m *vs.* 2.36 m, gap + 0.32), indicating that yaw is not the primary driver. Removing detector noise also eliminates the gap (3.09 m *vs.* 3.12 m, $p = 0.97$), suggesting that detector noise interacts with reduced viewing diversity. When only GPS bias is active, the imbalance partially reappears, although the paired difference remains above the conventional significance threshold (2.69 m *vs.* 2.48 m, $p = 0.067$). Overall, the diagnostic results support per-drone GPS bias as a necessary factor in the observed 30-seed imbalance, which is amplified by reduced viewing diversity and interacts with measurement noise. Under multiple-comparison (Holm) correction across the robustness family, and in a 300-seed re-run, the left-offset paired difference is not statistically significant ($p = 0.69$ at $n = 300$); we therefore present this analysis as a mechanism diagnosis rather than a claimed limitation.

Table 7. Robustness analysis (mean error in meters; W/L = wins/losses of the proposed selector *vs.* full-pool RANSAC). Bold indicates the lowest mean error; W/L counts may disagree with mean rankings when losses have a larger magnitude than wins.

Condition	Value	Proposed	Full-Pool	W/L
Formation	Line	**2.31**	2.49	14/16
Formation	V	**2.34**	2.44	11/19
Formation	Triangle	2.54	**2.47**	15/15
Target offset	Center	**2.34**	2.44	11/19
Target offset	Near	2.48	**2.44**	13/17
Target offset	Far	**2.33**	2.74	17/13
Target offset	Left	3.27	**2.92**	10/20
Target offset	Right	2.36	**2.34**	19/11

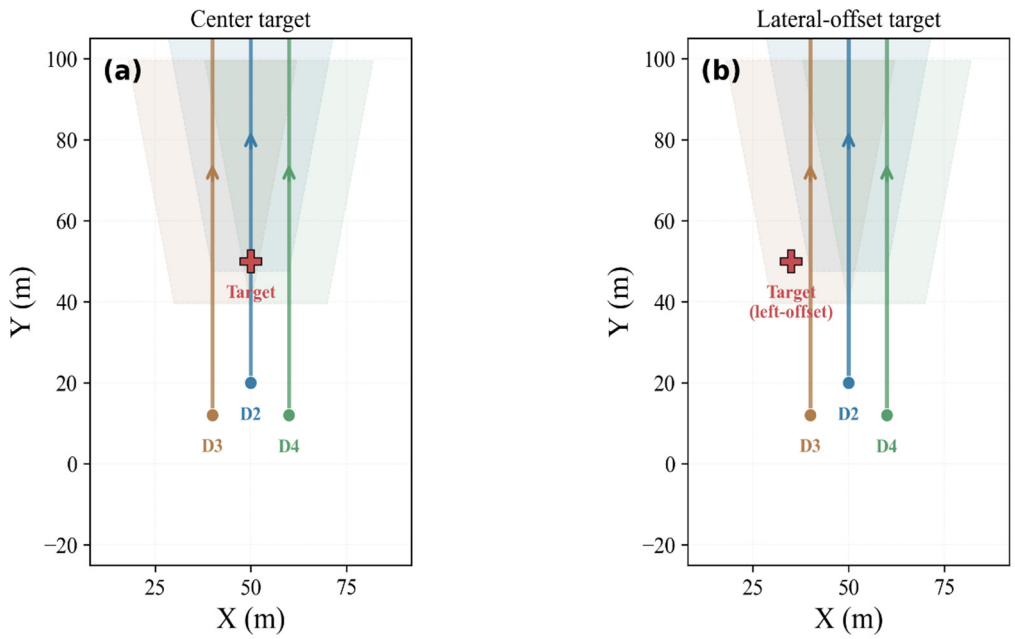

Figure 5. (**a**) Center versus left-offset target geometry. (**b**) The lateral offset reduces cross-drone balance by underrepresenting the far drone in the selected measurements.

Table 8. Lateral-offset diagnostic study (30 seeds, left-offset target). Gap = Proposed − Full-pool RANSAC; positive values indicate higher error for the proposed selector.

Condition	σ_{GPS}	σ_{ψ}	Proposed	Full-Pool	Gap
Baseline (all noise)	3.0	5°	3.27	2.92	+0.35
No GPS	0.0	5°	1.35	1.38	−0.03
No yaw	3.0	0°	2.68	2.36	+0.32
No detector noise	3.0	5°	3.09	3.12	−0.03
GPS only	3.0	0°	2.69	2.48	+0.21

5. Real-Flight Evaluation

To complement the simulation study, we conducted three flights with a low-cost quadrotor carrying a rigidly mounted, intrinsically calibrated monocular camera (focal length 1407 px, calibrated principal point; mount pitch −55° ± 2° measured statically) and an onboard ARM companion computer running the detection and logging stack (YOLOv8s, person class, confidence threshold 0.3). Flight 1 was at low altitude at dusk (~6 m), flight 2 added temporally separated passes (6–9 m), and flight 3 reached a higher altitude approaching the simulated regime (33–41 m). Ground truth is GPS-averaged (2-min static occupations of a fixed ground mark) and verified by manual annotation of the frames; no RTK reference was available, a limitation we return to below.

Per-drone GPS bias. The core assumption of this paper (a bias-dominant, slowly drifting per-drone GPS error) was tested in three independent ways: the same physical mark, re-occupied, appears displaced by 1.86 m after 9.6 min ($\chi^2 = 15.4$ against short-term scatter, with autocorrelation-corrected effective sample sizes), by 2.2 m between passes nine minutes apart (estimated on a static object), and by 3.47 m after 63 min ($\chi^2 = 42$; Figure 6). Manual annotation of the frames (the mark occupies the same pixel, ±2 px, in same-frame comparisons) rules out physical displacement. Beyond supporting the noise model, this drift motivates a pseudo-swarm protocol: passes separated by several minutes carry different bias realizations and are used here as stand-ins for distinct drones, enabling multi-view evaluation with a single vehicle. Consistent per-pass shifts of 1–2 m were likewise observed in flight 2 (8–19 min gaps).

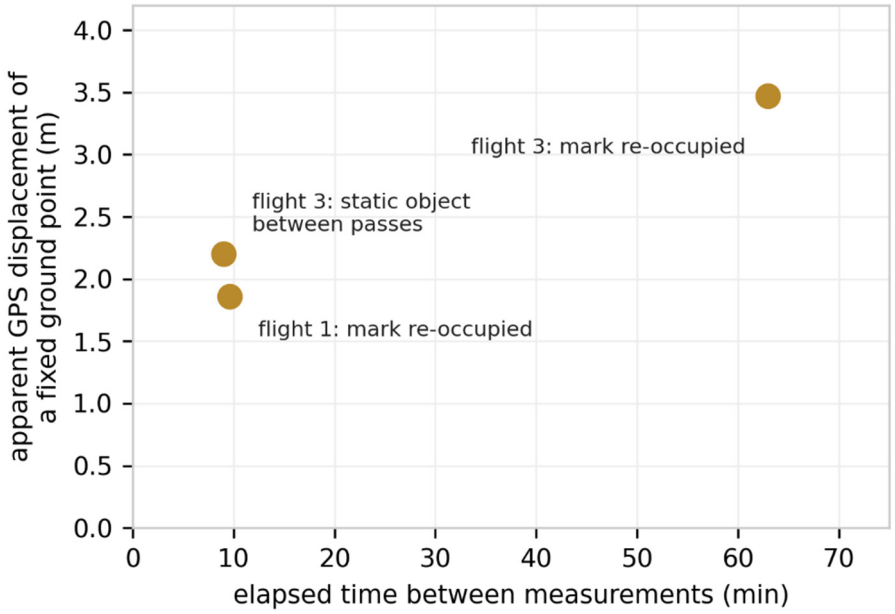

Figure 6. GPS bias decorrelates over time: the same fixed ground point, measured again after the indicated interval, appears displaced because the receiver bias changed between measurements (flights 1 and 3).

End-to-end accuracy and embedded runtime. In flight 1 (6 m), the full pipeline localizes the target to 1.26 m (selector) *vs.* 1.75 m (full pool). Both figures are measured against a GPS reference taken with the same receiver minutes before and after the flight, so the shared short-term bias largely cancels, and the comparison isolates detection and geometry errors. In flight 2, the winner alternates across passes, consistent with the accuracy-equivalence claim. On the onboard ARM computer, the selector cuts fusion time by a factor of 2.2 in 1448 real measurements (4.7 s *vs.* 10.2 s), which is smaller than the 4.1× measured on x86, but the advantage persists on embedded hardware.

A real-world failure mode: consensus capture by static false positives. In flight 3 (33–41 m), static person-like objects (an equipment case, traffic cones) were detected as "person" with median confidence 0.48 (*vs.* 0.65 for the true target) and, being static, accumulated roughly twice as many sightings. The RANSAC consensus was captured: the full-pool estimate lands 0.3 m from the distractor cluster and about 6 m from the operator (Figure 7). To measure achievable accuracy without circular reasoning, we blindly associate rays with targets (per-pass density clustering; the operator's manually annotated position is used only to label clusters post hoc). With correct associations, accuracy is 1.8–3.0 m at 33–41 m. The proposed selector recovers the correct target in the pass where target and distractor presence are comparable (2.3 m error, 9–10/10 RANSAC seeds, while full-pool RANSAC remains captured), but fails when false positives dominate 3:1. No confidence- or size-based cue can rescue a false-positive majority, since detector-selected false positives are person-sized by construction. This bounds unattended operation and motivates appearance-based identity cues as future work.

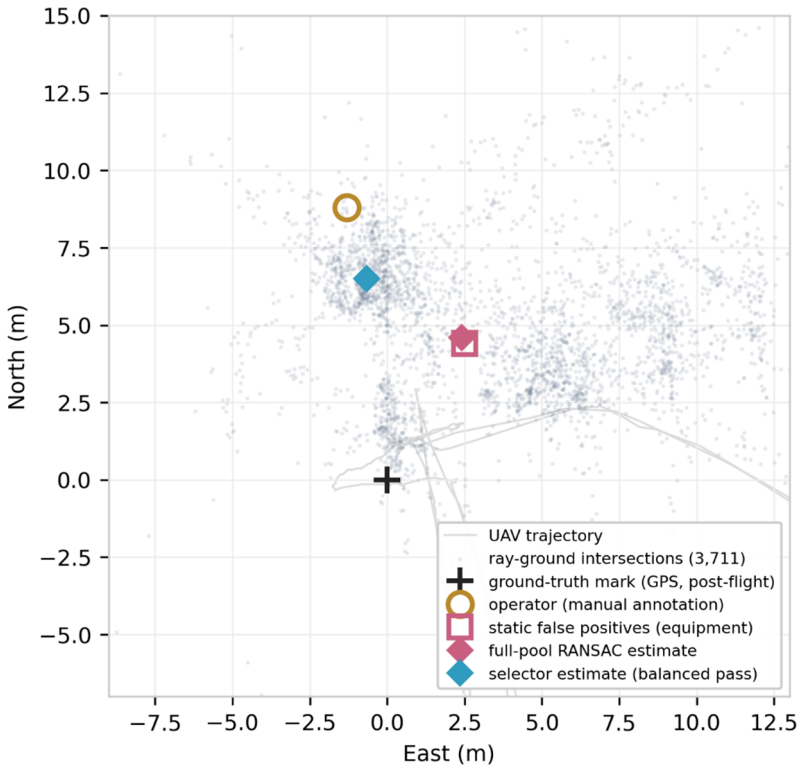

Figure 7. Overhead view of flight 3 (33–41 m): 3711 ray–ground intersections form two clusters, one around the operator (the true target) and one around the static false positives. The full-pool RANSAC estimate lands on the false-positive cluster; the selector estimate for the balanced pass lands near the operator. The ground-truth mark is the GPS reference of the coordinate frame, not the target.

Body-attitude effects. With a rigidly mounted camera (no gimbal), uncompensated body tilt during maneuvers shifts ray–ground intersections along the flight heading: the correlation between logged body pitch and along-heading residual is $\rho = -0.24$ ($p < 10^{-35}$, $n = 2801$), with a median +1.9 m shift under strong nose-down attitude (Figure 8). Compensating rays with the autopilot's logged attitude tightens the single-target cluster by 12–25%. These flights were gentle (84% of rays at |pitch| < 4°); in aggressive flight, this term would dominate, so the pipeline supports attitude compensation, applied here in post-processing.

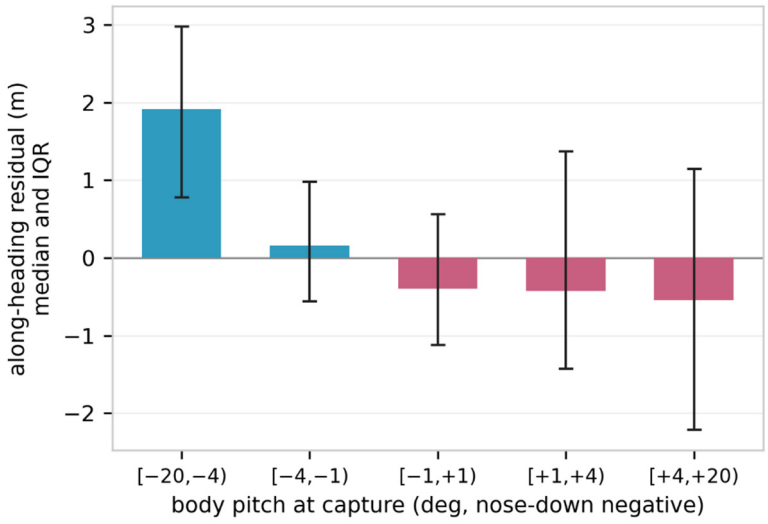

Figure 8. Uncompensated body attitude shifts ray impacts along the heading: median and interquartile range of the along-heading residual per body-pitch bin.

Model checks and limitations of the flights. The confidence-to-noise mapping is supported qualitatively (detections with confidence below 0.5 are ~60% noisier in pixel bearing), but the exact 1/confidence form is not (the noise floor is flat above 0.5); method rankings are unchanged under an alternative VisDrone-derived model. Within a single two-minute static occupation, the mean GPS solution drifts by 0.3–0.4 m in the most settled occupations, versus 4 m or more during the first minutes after power-up, which supports the quasi-constant-bias assumption at flyover timescales once the receiver has settled. Not covered by these flights and stated as limitations: RTK-grade ground truth, time-synchronization stress tests, rolling-shutter effects, non-flat terrain, and true multi-vehicle asynchrony (the pseudo-swarm uses one vehicle); an in-flight joint extrinsic calibration attempt did not converge under the multi-object failure mode above, so the mount pitch retains its statically measured value.

6. Discussion

The per-drone information saturation established in Section 3 has a clear practical implication: once the GPS bias ceiling is reached, additional detections from the same biased drone provide diminishing returns for geolocation accuracy. For practitioners designing multi-drone SAR systems, this suggests that, under per-drone GPS bias, increasing cross-drone diversity with independent GPS receivers can be more informative than increasing the frame rate or extending the observation time of a single drone.

This principle is reflected in the proposed selector's behavior. Its angular diversity criterion implicitly encourages source diversification, filtering out near-parallel measurements from the same drone under near-nadir geometry. Empirically, the selector concentrates roughly 12 measurements across the three drones (Table 3), in contrast to confidence-only and standard FIM selectors, which assign all measurements to a single drone. In the main setting, the selector achieves accuracy comparable to full-pool RANSAC while running 4.1× faster, because RANSAC operates on a pre-selected subset of ~12 measurements rather than the full pool of ~882. This trade-off, however, is regime-dependent. At larger GPS-bias scales ($\sigma_{\mathrm{GPS}} \geq 5$ m), full-pool RANSAC achieves slightly lower mean error, although the paired differences are not statistically significant. A similar effect appears for laterally offset targets, where the geometry reduces viewing diversity. The selector is therefore best positioned as a computationally efficient pre-selection method for the tested main regime, rather than as a uniformly preferable alternative.

Relative to classical bearing-only sensor selection [9,16], which typically treats measurement errors as independent across nodes or platforms, the bias-aware FIM analysis models per-drone GPS bias as a nuisance parameter. This leads to the saturation behavior observed in the analysis and motivates cross-drone diversification in the proposed selector. The selector combines angular diversity and detection confidence as a measurement pre-selection step before RANSAC, rather than directly optimizing the bias-aware FIM or relying on a formal submodular approximation guarantee. The additional FIM comparison supports this interpretation: the angular-diversity proxy is effective in the main operating regime, while explicit bias-aware FIM selection may be preferable under stronger GPS bias, though no tested regime shows a statistically significant difference.

The following limitations apply. The simulation uses empirically calibrated noise models, and the real-flight evaluation in Section 5, while supporting the bias model and demonstrating feasibility, is limited to a single vehicle (pseudo-swarm), GPS-averaged ground truth, and gentle flight conditions. The confidence-to-noise model is a proxy: replacing the baseline rule with the VisDrone-derived rule increases the selector's mean error from 2.34 m to 2.56 m, still statistically comparable to full-pool RANSAC (2.56 m *vs.* 2.49 m, $p = 0.50$), and confidence-only selection remains worse; real-flight data support the mapping only qualitatively (Section 5). Yaw bias is not modeled in the FIM formulation, a design choice justified in Section 3; it is quantified empirically instead, and the yaw sweeps show modest variation with no statistically significant paired differences relative to full-pool RANSAC. Under directed outlier contamination (20–30% of measurements aimed at a distractor, carrying realistically reduced confidence),

subset selection is more vulnerable to consensus capture than full-pool RANSAC (13% *vs.* 0% of seeds captured). In the real flight, the roles were reversed: full-pool RANSAC was captured, while the selector recovered the target during the balanced pass. Neither approach is robust to a false-positive majority, which motivates appearance-based identity cues: visual features that distinguish the true target from person-like static objects. The current formulation assumes a static target, a known ground plane, and centralized access to all detections and vehicle poses; the perturbation tests quantify target-height mismatch but not full terrain-aware geolocation. Moving targets, communication constraints, and decentralized fusion are left for future work. These limitations bound the empirical claims to the regimes considered here; the methodological contribution, a pre-selection step for RANSAC fusion motivated by the bias-aware FIM analysis, holds across the tested noise calibrations and is consistent with the real-flight results.

7. Conclusions

We presented a view selection approach for multi-UAV bearing-only geolocation under per-drone GPS bias. The bias-aware Fisher Information analysis shows a per-drone saturation property that explains why standard FIM-based selection and confidence-only approaches can degrade under correlated bias: they concentrate measurements on a single drone.

Motivated by this analysis, the proposed greedy selector combines angular diversity and detection confidence to pre-filter measurements before RANSAC, preserving accuracy comparable to full-pool RANSAC while reducing median processing time by 4.1× in the main configuration. Diagnostic experiments show that lateral target offsets reduce viewing diversity, although the associated accuracy difference is not statistically significant under multiple-comparison correction. In practice, the selector works as a pre-selection step for the tested main regime, and full-pool RANSAC may be preferred at larger GPS-bias scales.

As a first step toward deployment, the real-flight evaluation in Section 5 supports the per-drone bias model on hardware and documents a multi-object failure mode that shows where the method cannot yet operate unattended. Future work should extend the approach toward moving targets, appearance-based target identity cues, non-flat terrain, decentralized fusion, and online trajectory adaptation guided by real-time bias-aware FIM feedback.

Statement of the Use of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the authors used AI-based language tools to assist with grammar and stylistic refinement. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Author Contributions

Conceptualization, S.Q.A. and M.E.; Methodology, S.Q.A. and B.O.; Software, S.Q.A., L.L. and T.L.; Validation, S.Q.A. and T.R.; Formal Analysis, S.Q.A. and B.O.; Investigation, S.Q.A.; Resources, M.E.; Data Curation, S.Q.A. and T.R.; Writing—Original Draft, S.Q.A.; Writing—Review & Editing, S.Q.A., L.L., T.R., T.L., B.O. and M.E.; Visualization, S.Q.A. and T.L.; Supervision, B.O. and M.E.; Project Administration, M.E.; Funding Acquisition, M.E. All authors have read and agreed to the published version of the manuscript.

Ethics Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The source code and simulation scripts used to reproduce the experiments, tables, and figures in this paper are publicly available at https://github.com/sebastianquispearias/multi-uav-bearing-only-geolocation (accessed on 19 May 2026).

Funding

This work was supported by the Air Force Office of Scientific Research (AFOSR) under grant FA9550-23-1-0136.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Sambolek S, Ivasic-Kos M. Person Detection and Geolocation Estimation in Drone Images. *SN Comput. Sci.* **2025**, *6*, 358. DOI:10.1007/s42979-025-03869-7
2. Jocher G, Chaurasia A, Qiu J. Ultralytics YOLOv8; Ultralytics, 2023. Available online: https://github.com/ultralytics/ultralytics (accessed on 19 May 2026).
3. Lampesberger H, Hermann E, Zeilinger M. Object Geolocalization Using Consumer-Grade Devices: Algorithms and Applications. In Proceedings of the 4th International Workshop on Camera Traps, AI, and Ecology (CamTraps 2024), Hagenberg, Austria, 5–6 September 2024; pp. 71–83. DOI:10.5281/zenodo.13740788
4. Doğançay K. Bearings-only target localization using total least squares. *Signal Process.* **2005**, *85*, 1695–1710. DOI:10.1016/j.sigpro.2005.03.007
5. Ranacher P, Brunauer R, Trutschnig W, Van der Spek S, Reich S. Why GPS makes distances bigger than they are. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 316–333. DOI:10.1080/13658816.2015.1086924
6. Opromolla R. Magnetometer Calibration for Small Unmanned Aerial Vehicles Using Cooperative Flight Data. *Sensors* **2020**, *20*, 538. DOI:10.3390/s20020538
7. Bishop AN, Fidan B, Anderson BDO, Doğançay K, Pathirana PN. Optimality analysis of sensor-target localization geometries. *Automatica* **2010**, *46*, 479–492. DOI:10.1016/j.automatica.2009.12.003
8. Shamaiah M, Banerjee S, Vikalo H. Greedy Sensor Selection: Leveraging Submodularity. In Proceedings of the IEEE Conference on Decision and Control (CDC), Atlanta, GA, USA, 15–17 December 2010; pp. 2572–2577. DOI:10.1109/CDC.2010.5717225
9. Kirchner MR, Hespanha JP, Garagić D. Heterogeneous Measurement Selection for Vehicle Tracking using Submodular Optimization. In Proceedings of the 2020 IEEE Aerospace Conference, Big Sky, MT, USA, 7–14 March 2020; pp. 1–10. DOI:10.1109/AERO47225.2020.9172788
10. Xiong K, Wu X, Duan A, Huang Y, Leng S, He J. Information-Optimal Formation Geometry Design for Multimodal UAV Cooperative Perception. *arXiv* **2025**, arXiv:2512.12923. DOI:10.48550/arXiv.2512.12923
11. Lee CW, Waslander SL. UncertaintyTrack: Exploiting Detection and Localization Uncertainty in Multi-Object Tracking. In Proceedings of 2024 IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, 13–17 May 2024; pp. 4946–4953. DOI:10.1109/ICRA57147.2024.10610458
12. Solano-Carrillo E, Sattler F, Alex A, Klein A, Costa BP, Rodriguez AB, et al. UTrack: Multi-object Tracking with Uncertain Detections. In Proceedings of the ECCV Workshops, Milan, Italy, 29 September–4 October 2024. DOI:10.1007/978-3-031-91585-7_14
13. Jung H, Kang S, Kim T, Kim H. ConfTrack: Kalman Filter-Based Multi-Person Tracking by Utilizing Confidence Score of Detection Box. In Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2024; pp. 6569–6578. DOI:10.1109/WACV57701.2024.00645

14. Paulin G, Sambolek S, Ivasic-Kos M. Application of raycast method for person geolocalization and distance determination using UAV images in Real-World land search and rescue scenarios. *Expert Syst. Appl.* **2024**, *237*, 121495. DOI:10.1016/j.eswa.2023.121495
15. Sciacchitano A, Mudrik L, Kragelund S, Kaminer I. Multi-UAV Trajectory Optimization for Bearing-Only Localization in GPS Denied Environments. *arXiv* **2026**, arXiv:2602.11116. DOI:10.48550/arXiv.2602.11116
16. Kaplan LM. Global node selection for localization in a distributed sensor network. *IEEE Trans. Aerosp. Electron. Syst.* **2006**, *42*, 113–135. DOI:10.1109/TAES.2006.1603409
17. Dogancay K. 3D Pseudolinear Target Motion Analysis From Angle Measurements. *IEEE Trans. Signal Process.* **2015**, *63*, 1570–1580. DOI:10.1109/TSP.2015.2399869
18. Nemhauser GL, Wolsey LA, Fisher ML. An analysis of approximations for maximizing submodular set functions—I. *Math. Program.* **1978**, *14*, 265–294. DOI:10.1007/BF01588971
19. Joshi S, Boyd S. Sensor Selection via Convex Optimization. *IEEE Trans. Signal Process.* **2009**, *57*, 451–462. DOI:10.1109/TSP.2008.2007095
20. Krause A, Singh A, Guestrin C. Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies. *J. Mach. Learn. Res.* **2008**, 9, 235–284. Available online: https://www.jmlr.org/papers/volume9/krause08a.html (accessed on 19 May 2026).
21. Lin B, Wu L, Niu Y. End-to-End Vision-Based Cooperative Target Geo-Localization for Multiple Micro UAVs. *J. Intell. Robot. Syst.* **2022**, *106*, 13. DOI:10.1007/s10846-022-01639-8
22. Sun P, Tong S, Qin K, Luo Z, Lin B, Shi M. Low-Cost Real-Time Remote Sensing and Geolocation of Moving Targets via Monocular Bearing-Only Micro UAVs. *Remote Sens.* **2025**, *17*, 3836. DOI:10.3390/rs17233836
23. Georgiadis D, Politi E, Solmaz S, Dimitrakopoulos G. A review of localization and sensing technologies for UAV swarms in SAR missions. *EURASIP J. Adv. Signal Process.* **2026**, *2026*, 8. DOI:10.1186/s13634-025-01269-w
24. Fischler MA, Bolles RC. Random sample consensus. *Commun. ACM* **1981**, *24*, 381–395. DOI:10.1145/358669.358692
25. Chum O, Matas J. Matching with PROSAC—Progressive Sample Consensus. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–25 June 2005; pp. 220–226. DOI:10.1109/CVPR.2005.221
26. Chum O, Matas J, Kittler J. Locally Optimized RANSAC. In Proceedings of the DAGM Symposium on Pattern Recognition, Magdeburg, Germany, 10–12 September 2003; pp. 236–243. DOI:10.1007/978-3-540-45243-0_31
27. Raguram R, Chum O, Pollefeys M, Matas J, Frahm JM. USAC: A Universal Framework for Random Sample Consensus. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2022–2038. DOI:10.1109/TPAMI.2012.257
28. Hartley R, Zisserman A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2004. DOI:10.1017/CBO9780511811685
29. Jiang B, Luo R, Mao J, Xiao T, Jiang Y. Acquisition of Localization Confidence for Accurate Object Detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018. DOI:10.1007/978-3-030-01264-9_48
30. Zhu P, Wen L, Du D, Bian X, Fan H, Hu Q, et al. Detection and Tracking Meet Drones Challenge. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 7380–7399. DOI:10.1109/TPAMI.2021.3119563
31. PUC-Rio GrADyS Project. GrADyS-Sim NextGen. Available online: https://github.com/Project-GrADyS/gradys-sim-nextgen (accessed on 19 May 2026).