

Information-Optimal Formation Geometry Design for Multimodal UAV Cooperative Perception

Kai Xiong, *Member, IEEE*, Xingyu Wu, Anna Duan, Yongjun Huang, *Member, IEEE*, Supeng Leng, *Senior Member, IEEE*, Jianhua He, *Senior Member, IEEE*

Abstract—The cooperative perception performance of unmanned aerial vehicle (UAV) swarms is governed by three-dimensional (3D) formation geometry, which dictates target observability, sensor complementarity, and communication quality. Despite this, existing literature largely treats formation geometry as a secondary byproduct of trajectory planning, neglecting its impact on multimodal sensing scenarios where heterogeneous payloads must be spatially arranged to exploit complementary perception strengths under communication interference and hardware budget constraints. To bridge this gap, this paper proposes an information-optimal framework that jointly optimizes UAV-sensor allocation, formation configuration, and flight control for multimodal cooperative perception. Specifically, a submodular greedy strategy grounded in Fisher Information Matrix (FIM) determinant maximization determines the optimal number of UAV-sensor allocation. Building upon this allocation, an equivalent formation transition strategy reconfigures UAV positions to enhance field-of-view (FOV) coverage and suppress inter-UAV communication interference. Furthermore, a Lyapunov-stable distributed flight controller leveraging logarithmic potential fields generates energy-efficient trajectories. Extensive simulations demonstrate that our formation-aware design achieves 25.0% improvement in FOV coverage, 104.2% enhancement in communication signal strength, and 84.7% reduction in control effort compared to conventional benchmarks.

Index Terms—UAV 3D Formation, Multimodal Cooperative Perception, Fisher Information Matrix, Field-of-View.

I. INTRODUCTION

Unmanned Aerial Vehicle (UAV) swarms have emerged as transformative platforms for cooperative perception. They offer unprecedented capabilities in cooperative target tracking, surveillance, and the active-sensing phase of search and rescue (SAR) operations where a prior target state estimate is available [1–3]. Unlike single-UAV systems or terrestrial sensor networks, aerial swarms provide enhanced spatial coverage, improved viewpoint diversity, and parallel multimodal data acquisition through collaborative sensing [4]. It is noted

to distinguish between two different operational paradigms in SAR tasks: the *blind search* paradigm, where no prior information about the target exists, and the *active tracking* paradigm, where a preliminary target state estimate has been obtained through initial detection (e.g., distress signals, thermal signatures, or prior sweeps) [5]. This paper focuses on the latter one, in which the UAV swarm has access to a predicted target state and seeks to optimize its formation geometry for the information gain of subsequent observations. Here, the active tracking paradigm is consistent with the predictor-corrector framework widely adopted in information-driven sensor management [6]. In this framework, formation planning is conditioned on the best available state estimate, rather than requiring exact target knowledge.

However, current active tracking literature concentrates on trajectory planning, resource allocation, and perception algorithms while treating the formation geometry of UAV swarms as a secondary implementation detail [7–10]. This oversight significantly impacts multimodal cooperative perception performance in scenarios where heterogeneous payloads (vision cameras and LiDAR) are arranged to exploit their complementary strengths. Specifically, cameras provide rich visual semantics but suffer from depth estimation limitations, while LiDAR delivers precise three-dimensional (3D) geometric information at the cost of higher communication overhead and energy consumption [11]. Accordingly, the formation geometry can govern field-of-view (FOV) coverage, heterogeneous sensor complementarity, and target observability. Without proper geometric allocations, these complementary sensing capabilities cannot be effectively leveraged, leading to redundant observations and coverage blind spots.

Moreover, the geometric formation optimization is also challenged by practical collaboration constraints. In SAR tasks, UAVs typically operate in proximity, creating significant communication interference in dense formations [12]. While increasing swarm size enhances perception quality, it can significantly raise communication load, computational burden, and hardware costs. This requires a fundamental trade-off between multimodal perception performance and resource efficiency. It cannot be resolved through trajectory optimization alone. Furthermore, existing formation strategies typically assume homogeneous UAV capabilities [5, 13], without considering the spatial arrangement requirements of heterogeneous sensor payloads, which is a critical limitation for multimodal perception systems.

To bridge this critical gap, we propose an information-optimal framework that jointly optimizes geometric allocation

K. Xiong, X. Wu, Y. Huang, and S. Leng are with School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China.

A. Duan is with AVIC Chengdu Aircraft Design & Research Institute, 610041, Chengdu, 610041, China.

J. He is with School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK.

The financial support of Sichuan Provincial Natural Science Foundation under Grant 2026NSFSC1431, National Natural Science Foundation of China (NSFC), Grant No.62201122, AVIC United Technology Center for Intelligent Decision-making and Coordinated Control Mechanism Model Research, and EU's Horizon Europe MSCA Programme under Grant 101086228.

and sensor heterogeneity through Fisher Information Matrix (FIM) determinant maximization. The proposed formation optimization is activated after an initial target estimate is obtained. By treating formation design as a primary optimization variable rather than a byproduct of trajectory planning, we achieve simultaneous improvements in coverage completeness, communication Signal-to-Interference-plus-Noise Ratio (SINR), and flight energy efficiency. The main contributions of this paper are as follows:

- We establish a novel information-optimal framework for 3D UAV swarm formation design that explicitly quantifies the interdependence between spatial geometry, heterogeneous sensor modalities, and cooperative perception performance. By leveraging FIM determinant maximization as the optimization criterion, our framework provides a principled approach to determine information-optimal formation allocations in versatile perception environments where conventional homogeneous formation strategies fail to exploit multimodal sensing synergies.
- We formulate and solve a formation transition problem that addresses the FOV-oriented formation allocation under practical operational constraints. The proposed optimization model uniquely balances sensing information gain against FOV coverage completeness, communication overhead, and sensor payload costs. It incorporates realistic constraints including directional FOV limitations, inter-UAV interference effects, and heterogeneous payload capabilities that are neglected in existing literature.
- We develop a Lyapunov-stable distributed flight controller for energy-efficient formation transitions during dynamic operations. The control scheme employs a logarithmic potential field to ensure smooth trajectories with bounded control authority. This design provably minimizes unnecessary maneuvers and energy consumption while guaranteeing asymptotic flight stability. Extensive simulations show that our integrated approach achieves significant improvements over benchmarks in localization accuracy, coverage robustness, and energy efficiency across both aerial and ground target-search scenarios.

The remainder is organized as follows: Sec. II reviews related work. Sec. III presents system model. Sec. IV details the optimization scheme. Sec. V presents simulations, and Sec. VI concludes the paper.

II. RELATED WORK

This section reviews existing work in several interconnected research domains that collectively inspire our information-optimal UAV formation perception framework.

A. Multimodal Cooperative Perception

Multimodal cooperative sensing has emerged as a critical advancement to overcome the inherent limitations of single-platform perception systems in terms of coverage, accuracy, and environmental adaptability [14, 15]. Within the multi-UAV cooperative perception, Nitesh *et al.* [16] conducted an analysis of multimodal sensor integration techniques, which enhance detection precision, operational range, and overall

system reliability. Tian *et al.* [17] introduced a framework employing multi-view geometry and feature fusion to enable collaborative 3D object detection across multiple UAVs. Similarly, Qiao *et al.* [18] proposed a dual-window cross-attention mechanism that integrates LiDAR and camera features into a unified Bird's-Eye View representation. The development of standardized datasets has further accelerated research progress, with U2UData [19] establishing a benchmark featuring LiDAR point clouds, RGB/depth images, environmental sensing measurements, and 3D annotations for investigating multi-UAV multimodal perception and communication strategies. Notably, the aforementioned approaches are data-driven, relying on deep neural networks trained on large annotated datasets. Their dependence on extensive training data and the associated inference overhead limits their applicability in data-scarce scenarios and on resource-constrained platforms.

B. UAV Formation Design

Constrained by sensor heterogeneity and geometric allocation, generating optimal UAV formation to enhance cooperative perception performance remains a pivotal challenge. In that, FIM analysis has been widely conducted for various perception techniques including time difference of arrival (TDOA), angle of arrival (AOA), received signal strength (RSS), and their hybrid frameworks. Panwar *et al.* [13] integrated time of arrival (TOA), RSS, and AOA measurements, deriving the Cramér–Rao Lower Bound (CRLB) expression for joint measurements to evaluate target localization accuracy. Hung *et al.* [20] proposed a heterogeneous sensing multi-UAV tracking system modeling hybrid measurement configurations incorporating image sensors, TOA, and AOA to design optimal UAV formation. Further, Sahu *et al.* [21] developed a unified optimization framework integrating Alternating Direction Method of Multipliers (ADMM) with Majorization-Minimization techniques, for optimal sensor placements. However, the aforementioned FIM optimization typically yields idealized geometric allocations, which often results in severe FOV blind spots and inter-UAV communication interference. Moreover, prior work cannot directly address the 3D placement of diverse sensors (cameras and LiDARs), failing to fuse the asymmetric information contributions along different spatial axes.

C. Learning-Based Swarm Approaches

Beyond the model-based methods reviewed above, recent research has explored learning-based paradigms for swarm coordination. Reinforcement learning (RL) has been increasingly adopted to derive adaptive multi-UAV formation and guidance policies in dynamic environments. For instance, Wu *et al.* [7] developed an attention-based MADDPG formation controller for hybrid UAVs, while Gavin *et al.* [8] applied multi-agent RL to drone guidance for cooperative triangulation. Such methods achieve strong adaptability through trial-and-error policy learning. However, they generally require extensive offline training and incur non-trivial online inference costs, which are difficult to satisfy under the stringent size, weight, and power (SWaP) constraints of onboard UAV processors. In

contrast to these directions, our framework adopts a model-based, information-optimal formation that requires no deep learning training and relies solely on a preliminary target state estimate. It yields near-optimal formations with low computational latency, and jointly optimizes sensing accuracy, communication quality, and control efficiency within a unified geometric design.

III. SYSTEM MODEL

This section presents our framework for UAV formation-based multimodal cooperative perception, structured around three sequential strategies. As shown in Fig. 1, upon successful target identification, the UAV-sensor allocation and formation optimization strategies are executed only once: the former determines the number of UAVs and their sensor categories (LiDAR or Camera), and the latter enhances FOV coverage while preserving perception and communication quality. This one-shot initialization defines the mathematically optimal geometric allocation, avoiding redundant high-overhead optimization during cooperative flight. In contrast, the flight control strategy operates continuously, iteratively updating the UAV kinematics in response to target maneuvers while maintaining the predefined optimal topology. This sequential approach holistically optimizes geometric allocation, sensor complementarity, communication, and flight energy.

A. System Assumption

The UAV formation for cooperative perception is demonstrated in Fig. 2, where heterogeneous UAVs collaboratively monitor an area for target detection and tracking. These UAVs are equipped with complementary sensing modalities, some with image sensors (cameras) and others with point cloud sensors (LiDAR).

The target, which may move on the ground or in the air, has a position vector $\mathcal{P}_{tar}(t) = [x_{tar}(t), y_{tar}(t), z_{tar}(t)]^T$ at time t . Each UAV i has a position $\mathcal{P}_i(t) = [x_i(t), y_i(t), z_i(t)]^T$. We assume zero pitch angles during steady flight, and define the heading vector as $\mathbf{b}_i = [\cos \theta_i, \sin \theta_i, 0]$, where θ_i denotes the yaw angle. Consequently, the relative 3D position vector is simply expressed as $\mathcal{P}_{i/tar}(t) = \mathcal{P}_i(t) - \mathcal{P}_{tar}(t)$, and its projection onto the horizontal X-Y plane is $\mathcal{P}_{i/tar}^{xy}(t) = [x_i(t) - x_{tar}(t), y_i(t) - y_{tar}(t), 0]^T$.

In practical SAR tasks, the target position $\mathcal{P}_{tar}(t)$ may not be directly observable. To address this, we formulate the cooperative perception as a predictor-corrector active sensing process. At planning step t , the UAV swarm employs a tracking filter (e.g., extended Kalman filter) to estimate the predicted target state $\hat{\mathcal{P}}_{tar}(t|t-1)$ based on previous observations. The formation geometry is then proactively optimized by relying on this prior state to acquire high-fidelity measurements, which subsequently refine the posterior state $\hat{\mathcal{P}}_{tar}(t|t)$. Subsequently, we will elaborate on the UAV-sensor allocation, i.e., the formation allocation, for the multimodal cooperative perception.

B. Multimodal Cooperative Perception

This subsection quantifies multimodal cooperative perception via the FIM metric. We analyze two complementary modalities (vision cameras and point-cloud LiDAR) deriving their measurement models, Jacobians, and FIM formulations, and explicitly modeling each sensor's information contribution within the formation-aware framework. The formulation supports optimal sensor allocation for localization accuracy while accounting for FOV limitations, measurement noise, and communication interference.

1) *Perception Measurement of Camera*: We construct the camera measurement model in three sequential steps: (i) transforming the target into the camera's local frame, (ii) applying the pinhole projection to obtain pixel coordinates, and (iii) computing the Jacobian for the FIM.

Here, the camera maps the 3D target position to the 2D image plane through a perspective projection. Building on this, we transform the global coordinates into the camera's local coordinate system. The pixel coordinates of the projection are denoted by $g_{cam,i} = [\mu_i, \nu_i]^T$, which is given as:

$$g_{cam,i} = h_{cam,i}(\mathcal{P}_{tar}, \mathcal{P}_i, \mathbf{b}_i) + \epsilon_{cam}, \quad (1)$$

where ϵ_{cam} represents the measurement noise vector associated with the resolution level of the camera sensor and communication interference. $\mathbf{b}_i$ is the heading vector defined in Sec. III-A, which is determined by the UAV's yaw angle θ_i .

$h_{cam,i}(\cdot)$ is the projection function mapping the target position to the camera coordinate system. Assuming the camera's optical axis is aligned with the UAV's yaw angle θ_i and the pitch angle is zero. Hence, the target's coordinates in the camera's local frame $(X_c, Y_c, Z_c)^T$ are obtained via the rotation matrix R_g^c :

$$\begin{aligned} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} &= R_g^c(\mathcal{P}_i - \mathcal{P}_{tar}) \\ &= \begin{bmatrix} -\sin \theta_i & \cos \theta_i & 0 \\ 0 & 0 & -1 \\ \cos \theta_i & \sin \theta_i & 0 \end{bmatrix} \begin{bmatrix} x_i - x_{tar} \\ y_i - y_{tar} \\ z_i - z_{tar} \end{bmatrix}, \end{aligned} \quad (2)$$

where $Z_i = \cos \theta_i(x_i - x_{tar}) + \sin \theta_i(y_i - y_{tar})$ denotes the target depth along the camera optical axis.

Based on the standard pinhole camera model, the forward projection function $\pi(\cdot)$ transforms this 3D point to the pixel coordinates:

$$\begin{aligned} h_{cam,i}(\mathcal{P}_{tar}, \mathcal{P}_i, \mathbf{b}_i) &= \begin{bmatrix} \mu_i \\ \nu_i \end{bmatrix} = \begin{bmatrix} f_x \frac{X_c}{Z_i} + c_x \\ f_y \frac{Y_c}{Z_i} + c_y \end{bmatrix} \\ &= \begin{bmatrix} -f_x \frac{\cos \theta_i(y_i - y_{tar}) - \sin \theta_i(x_i - x_{tar})}{\cos \theta_i(x_i - x_{tar}) + \sin \theta_i(y_i - y_{tar})} + c_x \\ -f_y \frac{z_i - z_{tar}}{\cos \theta_i(x_i - x_{tar}) + \sin \theta_i(y_i - y_{tar})} + c_y \end{bmatrix}, \end{aligned} \quad (3)$$

where f_x and f_y are the focal lengths, and c_x and c_y are the principal point offsets.

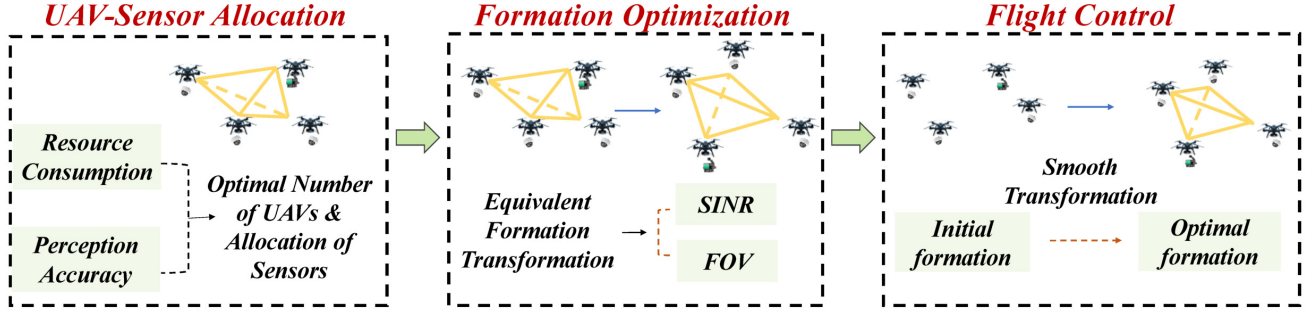


Fig. 1: The illustration of the comprehensive framework for the cooperative perception of UAV swarms.

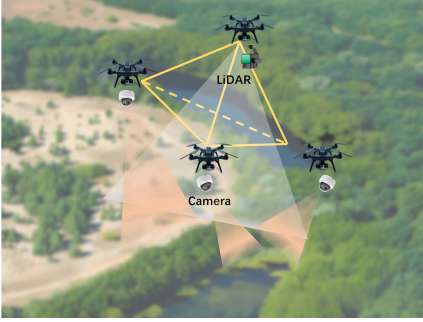


Fig. 2: UAV swarm-based SAR scenario.

Applying the chain rule to $h_{\text{cam},i}$ with respect to the target state yields the 2×3 Jacobian:

$$\begin{aligned} \mathbf{O}_{\text{cam},i} &= \frac{\partial h_{\text{cam}}}{\partial \mathcal{P}_{\text{tar}}} \\ &= \begin{bmatrix} \frac{-f_x(y_i - y_{\text{tar}})}{Z_i^2} & \frac{f_x(x_i - x_{\text{tar}})}{Z_i^2} & 0 \\ \frac{-f_y \cos \theta_i(z_i - z_{\text{tar}})}{Z_i^2} & \frac{-f_y \sin \theta_i(z_i - z_{\text{tar}})}{Z_i^2} & \frac{f_y}{Z_i} \end{bmatrix}. \end{aligned} \quad (4)$$

Therefore, the FIM of the i -th UAV with a camera sensor is defined as:

$$\mathcal{F}_{\text{cam},i} = \mathbf{O}_{\text{cam},i}^T \mathbf{Q}_{\text{cam}}^{-1} \mathbf{O}_{\text{cam},i}, \quad (5)$$

where $\mathbf{Q}_{\text{cam},i}$ denotes the measurement covariance matrix of the camera sensor, which depends on the pixel resolution and the communication SINR.

2) *Perception Measurement of LiDAR:* The LiDAR measurement comprises range, azimuth, and elevation, stacked into the full measurement vector and Jacobian. The range measurement is:

$$d_i = h_d(\mathcal{P}_{\text{tar}}, \mathcal{P}_i) + \varepsilon_d, \quad (6)$$

where ε_d is the measurement noise and $h_d(\mathcal{P}_{\text{tar}}, \mathcal{P}_i)$ is the distance between the UAV i and target:

$$h_d(\mathcal{P}_{\text{tar}}, \mathcal{P}_i) = \sqrt{(x_i - x_{\text{tar}})^2 + (y_i - y_{\text{tar}})^2 + (z_i - z_{\text{tar}})^2}. \quad (7)$$

Regarding 3D angle measurement, it is composed of the azimuth angle and the elevation angle of the target. The azimuth angle measurement β is expressed as:

$$\beta_i = h_A(\mathcal{P}_{\text{tar}}, \mathcal{P}_i) + \varepsilon_A, \quad (8)$$

where ε_A is the measurement noise and $h_A(\mathcal{P}_{\text{tar}}, \mathcal{P}_i)$ represents the azimuth angle difference, which is:

$$h_A(\mathcal{P}_{\text{tar}}, \mathcal{P}_i) = \tan^{-1} \left(\frac{y_i - y_{\text{tar}}}{x_i - x_{\text{tar}}} \right). \quad (9)$$

In contrast, the elevation angle measurement δ is given as:

$$\delta_i = h_p(\mathcal{P}_{\text{tar}}, \mathcal{P}_i) + \varepsilon_p, \quad (10)$$

where ε_p is the noise and $h_p(\mathcal{P}_{\text{tar}}, \mathcal{P}_i)$ represents the elevation angle difference, which is:

$$h_p(\mathcal{P}_{\text{tar}}, \mathcal{P}_i) = \tan^{-1} \left(\frac{z_i - z_{\text{tar}}}{\sqrt{(x_i - x_{\text{tar}})^2 + (y_i - y_{\text{tar}})^2}} \right). \quad (11)$$

Based on the above analysis, the measurement of the LiDAR h_{lidar} is composed by:

$$h_{\text{lidar},i}(\mathcal{P}_{\text{tar}}, \mathcal{P}_i) = \begin{bmatrix} h_d(\mathcal{P}_{\text{tar}}, \mathcal{P}_i) \\ h_A(\mathcal{P}_{\text{tar}}, \mathcal{P}_i) \\ h_p(\mathcal{P}_{\text{tar}}, \mathcal{P}_i) \end{bmatrix}. \quad (12)$$

Hereafter, the Jacobian matrix $\mathbf{O}_{\text{lidar},i}$ of the LiDAR measurement for the i -th UAV is:

$$\begin{aligned} \mathbf{O}_{\text{lidar},i} &= \frac{\partial h_{\text{lidar}}}{\partial \mathcal{P}_{\text{tar}}} \\ &= \begin{bmatrix} -\frac{x_i - x_{\text{tar}}}{d_i} & -\frac{y_i - y_{\text{tar}}}{d_i} & -\frac{z_i - z_{\text{tar}}}{d_i} \\ \frac{\sin \beta_i}{d_i^{xy}} & -\frac{\cos \beta_i}{d_i^{xy}} & 0 \\ \frac{(z_i - z_{\text{tar}}) \cos \beta_i}{d_i^2} & \frac{(z_i - z_{\text{tar}}) \sin \beta_i}{d_i^2} & -\frac{d_i^{xy}}{d_i^2} \end{bmatrix}, \end{aligned} \quad (13)$$

where d_i^{xy} is the horizontal distance of the i -th UAV projected onto the tracking target, which is expressed as:

$$d_i^{xy} = \sqrt{(x_i - x_{\text{tar}})^2 + (y_i - y_{\text{tar}})^2}. \quad (14)$$

Finally, the FIM of LiDAR for the i -th UAV is:

$$\mathcal{F}_{\text{lidar},i} = \mathbf{O}_{\text{lidar},i}^T \mathbf{Q}_{\text{lidar}}^{-1} \mathbf{O}_{\text{lidar},i}, \quad (15)$$

where $\mathbf{Q}_{\text{lidar}}$ is the measurement covariance matrix, which is related to the measurement and wireless transmission noise. Subsequently, we elaborate on the resource constraints of the communication interference, flight energy, and sensor budgets.

As shown in Eq. (4) and Eq. (13), the Jacobian matrices $\mathbf{O}_{\text{cam},i}$ and $\mathbf{O}_{\text{lidar},i}$ inherently depend on the unknown true

target position $\mathcal{P}_{tar}$. In our active sensing framework, these Jacobians are evaluated at the predicted state $\hat{\mathcal{P}}_{tar}(t|t-1)$ rather than the true state. This local approximation is a standard practice in information-driven sensor placement, enabling the swarm to maximize the expected information gain upon the current best estimate of the target. We note that this design choice explicitly scopes our framework to the active tracking regime. In the blind search phase, where no prior estimate exists, alternative exploration strategies (e.g., coverage-based or ergodic search) are required to establish initial target detection before our formation optimization can be applied.

For mutually independent measurements, the total FIM is equal to the sum of the individual FIMs. Assuming the formation includes N_c camera-equipped UAVs and N_l LiDAR-equipped UAVs, and the sensor noises are mutually independent, the global FIM with respect to the target position, $\mathcal{F}_{total}$, can be expressed as:

$$\mathcal{F}_{total} = \sum_{i=1}^{N_c} \mathcal{F}_{cam, i} + \sum_{j=1}^{N_l} \mathcal{F}_{lidar, j}. \quad (16)$$

In practical deployment, our framework operates in the active tracking paradigm, assuming a preliminary target state is available via initial detection. Should this prediction deviate significantly, the swarm reverts to coverage-based search to re-acquire the target. To ensure mathematical tractability, we assume zero UAV pitch angles during steady-state tracking and an unobstructed line-of-sight (LoS). The LoS assumption holds well for aerial targets and open-terrain tracking, while terrain- and vegetation-induced occlusions are accommodated as model extensions in Sec. IV-B.

The diagonal structure of the measurement covariance matrices ($\mathbf{Q}_{cam}$ and $\mathbf{Q}_{lidar}$) is justified by statistical independence at two levels: within each UAV, the measurement channels of a given sensor (the pixel coordinates (μ, ν) of the camera, and the range, azimuth, and elevation of the LiDAR) are driven by separate electronic processes. Across UAVs, the heterogeneous sensors are mounted on physically isolated airframes. The former underpins the FIM invariance under the equivalent transition in Sec. IV-B, while the latter ensures the additivity of per-UAV FIMs. Multimodal data fusion additionally relies on tight time synchronization, precise camera-LiDAR extrinsic calibration, and reliable cross-view data association. Here, data association is facilitated by the complementarity of visual semantics and geometric depth, and is further strengthened by the FOV-maximizing nature of our optimization, which can minimize occlusion-induced association failures.

These idealized baselines are reasonable in practice: the compact size of the UAV swarm renders air-to-air propagation delays negligible relative to the sensing intervals. Minor operational deviations, such as temporary pitch fluctuations or synchronization delays, can be modeled as zero-mean additive perturbations that inflate the diagonal entries of $\mathbf{Q}$. This degrades the achievable Cramér–Rao bound but preserves the FIM invariance property, as quantitatively validated in Sec. V. More substantial violations, such as persistent strongly correlated noise, lie beyond the current scope.

C. Resource Consumption Analysis

In the multimodal cooperative perception of the UAV swarm, the resource consumption mainly includes three aspects: communication consumption, flight maneuver energy consumption, and the inherent costs of heterogeneous sensors. Quantitative analysis of these costs is crucial for evaluating the feasibility of the system and optimizing the overall efficiency.

1) *SINR*: In realistic dense UAV swarms, the air-to-air communication links suffer not only from distance-dependent attenuation but also from environmental shadowing and multi-path fading. To reflect harsher communication conditions, the received signal power at UAV j from UAV i is modeled as:

$$p_{i,j}^{rec} = P_{tx} \rho_0 \|\mathcal{P}_j - \mathcal{P}_i\|^{-\alpha} |h_{i,j}|^2, \quad (17)$$

where P_{tx} is the uniform transmission power, α is the path loss exponent, and ρ_0 is the reference path loss at 1 meter [22]. Crucially, $|h_{i,j}|^2$ is the channel gain coefficient encompassing small-scale fading (e.g., Rayleigh or Rician fading) and log-normal shadowing, which introduces stochastic fluctuations into the link quality.

Consequently, the SINR of the communication link between the i -th and j -th UAV is expressed as:

$$\eta_{ij} = \frac{p_{i,j}^{rec}}{\sum_{k \in U, k \neq i,j} p_{k,j}^{rec} + \sigma^2}, \quad (18)$$

where σ^2 is the background noise power, U is the set of UAVs engaged in cooperative sensing, and the denominator accounts for the severe co-channel interference generated by other closely operating UAVs in the swarm.

2) *Time-Frequency Communication Resource*: Regarding the i -th UAV, the time-frequency resource block for transmitting sensing data can be expressed as:

$$\mathcal{O}_i^{comm} = B_i \cdot T_i, \quad (19)$$

where B_i is the occupied bandwidth for the i -th UAV's data transmission. T_i represents the corresponding transmission duration for the i -th UAV's, which can be regarded as the wireless channel occupation time. Due to the inherently larger data volume of LiDAR-based point clouds compared to RGB images, LiDAR-equipped UAVs generally incur higher time-frequency resource block consumption with the prescribed transmission rate.

3) *Sensor Hardware Budgets*: Multimodal sensors incur distinct operational and manufacturing budgets. The hardware cost C^{sensor} of onboard UAV sensors can generally be considered as a constant. Additionally, the hardware cost C_i^{lidar} of LiDAR is higher than that of cameras C_i^{cam} . Thus, $C_i^{lidar} > C_i^{cam}$.

D. Properties of FIM

To facilitate the analysis, feasible UAV deployment positions are modeled as a finite discrete set V , from which a subset S is selected to form the active formation configuration. The objective function $f(S)$ evaluates this selected subset by maximizing the log-determinant of its FIM. To guarantee an optimal greedy solution, $f(S)$ satisfies two fundamental properties [23]: *monotonicity* ($f(T) \geq f(S)$ for all $S \subseteq T \subseteq V$)

and *submodularity* ($f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)$ for any $v \in V \setminus T$). Monotonicity implies that adding a UAV never degrades perception, while submodularity encapsulates the diminishing marginal returns of adding new sensors. Under a fixed cardinality constraint, these properties guarantee that a greedy strategy yields a lower-bound performance of $(1 - 1/e) \approx 63.2\%$ relative to the global optimum.

IV. FORMATION AWARE STRATEGIES

To address the cooperative perception problem, including sensing performance, communication, and flight energy constraint, we divide the UAV formation scheme into three subproblems: the UAV-sensor allocation, FOV-oriented formation optimization, and energy-efficient flight control problems.

A. UAV-Sensor Allocation

The first subproblem aims to determine the optimal number of UAVs and the multimodal sensor allocation (camera and LiDAR) within the swarm. The objective is to maximize the collaborative sensing accuracy for target tracking, while considering the diminishing reward of new node incorporation and the associated costs of communication overhead and hardware deployment.

We adopt the D-optimality (Determinant-optimality) criterion from optimal experimental design, which leads to the following objective function [24]:

$$f(U) = \log \det \mathcal{F}(U), \quad (20)$$

where $U \subseteq V$ is the subset of the candidate set V . $\mathcal{F}(U)$ is the FIM for the selected allocation. Crucially, the log-determinant of the FIM has been proven to satisfy both monotonicity and submodularity [24, 25], thereby validating the theoretical applicability of the proposed greedy selection strategy. To balance sensor performance of different UAVs with practical resource constraints, we augment the objective with penalty terms for communication time-frequency resource and hardware costs:

$$\mathcal{J}(U) = \log \det \mathcal{F}(U) - w_1 \sum_{v \in U} \mathcal{E}_v^{\text{comm}} - w_2 \sum_{v \in U} C_v^{\text{sensor}}, \quad (21)$$

where $\mathcal{E}_v^{\text{comm}}$ denotes the time-frequency resources block of UAV v , C_v^{sensor} is the sensor hardware costs, and w_1, w_2 are weight coefficients that regulate the trade-off between perception performance and resource utilization. Based on the above analysis, we provide the optimization model for formation cooperative perception as:

$$\begin{aligned} (\text{P1}): \quad & \max_{U \subseteq V} \mathcal{J}(U) \\ \text{s.t.} \quad & C_1 : |U| \leq N_{\max} \\ & C_2 : d_{\min} \leq d_v \leq d_{\max}, \quad \forall v \in U \\ & C_3 : \beta_{\min} \leq \beta_v \leq \beta_{\max}, \quad \forall v \in U \\ & C_4 : \delta_{\min} \leq \delta_v \leq \delta_{\max}, \quad \forall v \in U \\ & C_5 : s_v \in \{\text{Camera, LiDAR}\}, \quad \forall v \in U \end{aligned} \quad (22)$$

in which C_1 is the maximum number of deployed UAVs according to practical conditions and coordination capabilities. C_2 gives the distance range constraints, designed to ensure a safe separation from the target while simultaneously keeping the UAV within the sensor's effective perception coverage. C_3 and C_4 specify the azimuth and elevation angle constraints, respectively, guaranteeing that the target remains within the sensor's effective FOV. And C_5 represents the available sensor category, i.e., the camera and LiDAR.

Notably, the monotonicity and submodularity properties of the FIM enable us to employ efficient greedy selection algorithms with theoretical optimal value guarantees. The marginal gain of incorporating an additional UAV v into an existing set U_j that contains j number of UAVs is given as:

$$\Delta(v|U_j) = f(U_j \cup v) - f(U_j). \quad (23)$$

Moreover, to implement the greedy selection algorithm, we introduce a utility function that incorporates both marginal gains and system costs:

$$G(v) = \Delta(v|U_j) - w_1 \mathcal{E}_v^{\text{comm}} - w_2 C_v^{\text{sensor}}. \quad (24)$$

The core of our greedy selection strategy involves iteratively choosing UAV spatial allocations from a finite candidate set V , constructed through discretization of the solution space encompassing all feasible UAV positions and sensor combinations. Specifically, the candidate set V is given as:

Spatial Discretization: We model the airspace around the target in a target-centered spherical coordinate system, discretizing the radial distance d at intervals Δd within $[d_{\min}, d_{\max}]$ (bounded by the effective sensing range) and uniformly sampling the azimuth β and elevation δ at resolutions $\Delta\beta$ and $\Delta\delta$. It yields a grid of candidate points $p_{\text{candidate}} = (d, \beta, \delta)$.

Sensor Assignment: For each candidate position, we generate two configurations—one camera and one LiDAR—which doubles the candidate set but enables optimal modality selection under information and hardware constraints.

View Direction: The yaw θ_i is set by aligning the sensor's optical axis toward the target, making it a deterministic function of the relative position rather than an independent variable, which reduces the search dimensionality.

Algorithm 1 summarizes the complete greedy selection procedure. The algorithm terminates when the maximum utility gain $G(v_t)$ falls below a predefined threshold ρ , indicating that the inclusion of additional UAVs no longer offsets their associated communication and hardware costs. We note that the classical $(1 - 1/e)$ -approximation guarantee formally applies to greedy maximization of a monotone submodular function under a cardinality constraint. The cost-penalized utility $G(v)$ and threshold-based stopping rule in our formulation depart from this standard setting, so the formal bound does not directly extend to the full algorithm. Nevertheless, since the underlying FIM objective $f(U)$ retains monotonicity and submodularity, the $(1 - 1/e)$ bound on $f(U)$ remains a meaningful performance reference.

More importantly, the computational complexity of this procedure is rigorously bounded by $\mathcal{O}(M|V|)$, where M denotes the final number of selected UAVs and $|V|$ is the

size of the discrete candidate set. In each of the M iterations, the algorithm evaluates the marginal gain for the remaining candidate allocations. Given that the practical swarm size M is typically small, the algorithm effectively scales linearly with the search space size. This manageable linear complexity makes the proposed greedy strategy well suited for real-time onboard execution on resource-constrained UAV platforms in time-critical SAR tasks.

Algorithm 1: UAV-sensor Greedy Selection

```

1 Input: Candidate set  $V$ , cost coefficients  $w_1, w_2$ ,
   time-frequency resource  $\mathcal{E}$ , hardware budgets  $C_v^{sensor}$ 
2 Output: UAV-sensor allocation  $U_M$ 
3  $U_0 \leftarrow \emptyset$ ;  $M \leftarrow 0$ 
4 while true do
5    $M \leftarrow M + 1$ 
6   for each  $v \in V \setminus U_{M-1}$  do
7      $\Delta(v|U_{M-1}) \leftarrow$ 
        $\log \det \mathcal{F}(U_{M-1} \cup v) - \log \det \mathcal{F}(U_{M-1})$ 
8      $G(v) \leftarrow \Delta(v|U_{M-1}) - w_1 \mathcal{E}_v^{comm} - w_2 C_v^{sensor}$ 
9    $v_t \leftarrow \arg \max_{v \in V \setminus U_{M-1}} G(v)$ 
10  if  $G(v_t) \leq \text{termination threshold } \rho$  then
11    break
12   $U_M \leftarrow U_{M-1} \cup \{v_t\}$ 

```

B. Formation Optimization

Upon the UAV-sensor allocation through Alg. 1, we now address the optimal spatial formation to maximize the FOV coverage without FIM loss. While the selected UAVs and sensors provide theoretically optimal target localization accuracy, their initial geometric allocation may exhibit significant FOV gaps and communication interference that compromise operational robustness in dynamic SAR scenarios. This subsection introduces an equivalent transition framework that preserves the fundamental perception accuracy (the FIM determinant) while strategically reconfiguring UAV positions to maximize FOV coverage completeness and communication quality.

To ensure continuous target tracking during surveillance operations, FOV coverage constitutes a critical performance metric beyond mere perception accuracy. An optimal formation must not only maintain target observability but also provide spatially balanced observation perspectives to mitigate occlusion risks and accommodate potential target maneuvers. We model each UAV's FOV coverage as a rectangular pyramidal frustum determined by three geometric parameters: maximum perception range $d_{\max}$, horizontal FOV (HFOV) γ , and vertical FOV (VFOV) κ . Crucially, $d_{\max}$ denotes the effective sensing range. It is physically constrained by the minimum resolution that the onboard sensors require relative to the target's physical dimensions. Thus, the selection of $d_{\max}$ implicitly accounts for the expected sizes of characteristic objects, such as humans or vehicles, within the specific SAR task context. The current FOV model assumes an unobstructed LoS; terrain- and vegetation-induced occlusions can be incorporated by introducing a probabilistic LoS coefficient into the

global FIM and by inflating the covariance $\mathbf{Q}$, which we leave as a model extension. Fig. 3 presents the pyramidal frustum-based FOV demonstration.

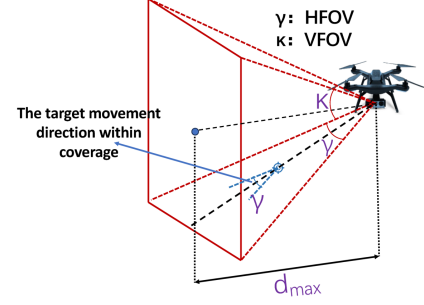


Fig. 3: Demonstration of FOV.

To capture fine-grained perception gains, we measure the coverage quality using a *quality-weighted coverage metric* that explicitly reflects viewing geometry and observation range. Then, we discretize the horizontal space surrounding the target into $\mathcal{N}$ equally spaced azimuth directions represented by unit vectors $\mathbf{n}_k = [\cos \phi_k, \sin \phi_k, 0]^T$, where $\phi_k = 2\pi k/\mathcal{N}$ for $k = 0, 1, \dots, \mathcal{N} - 1$. The number of sampled directions $\mathcal{N}$ is determined by the required angular resolution. Denote $\varphi_{i,k}$ the angular deviation between the discrete direction $\mathbf{n}_k$ and the i -th UAV's horizontal observation vector $\mathcal{P}_{i/tar}^{xy}$:

$$\varphi_{i,k} = \arccos \left(\frac{\mathbf{n}_k \cdot \mathcal{P}_{i/tar}^{xy}}{\|\mathcal{P}_{i/tar}^{xy}\|} \right). \quad (25)$$

In practical optical and LiDAR sensors, perception quality degrades gracefully as the target deviates from the optical center and as the relative distance increases. Thus, we define a continuous quality-weighted coverage intensity $q(i, k)$ contributed by UAV i to direction $\mathbf{n}_k$:

$$q(i, k) = \begin{cases} \exp \left(-\frac{\varphi_{i,k}^2}{2\sigma_\gamma^2} \right) \cdot \frac{d_{\text{ref}}^2}{\|\mathcal{P}_{i/tar}^{xy}\|^2}, & \text{if } \varphi_{i,k} \leq \frac{\gamma_i}{2} \\ & \text{and } \|\mathcal{P}_{i/tar}\| \leq d_{\max}, \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

where σ_γ controls the Gaussian sensitivity of perception degradation off the optical axis, and d_{ref} is a normalizing reference distance. Physically, this intensity metric models the sensor-specific detection reliability and spatial resolution capacity, explicitly capturing the confidence drop at the periphery of the FOV. For each discrete direction $\mathbf{n}_k$, the total coverage intensity is the superposition of contributions from all M deployed UAVs (obtained from Alg. 1):

$$\Phi(\mathbf{n}_k) = \sum_{i=1}^M q(i, k). \quad (27)$$

To explicitly penalize completely uncovered directions (blind spots) and promote coverage integrity, we introduce a penalty multiplier ξ :

$$\xi = 1 - \frac{\mathcal{N}_{\text{uncovered}}}{\mathcal{N}}, \quad (28)$$

where $\mathcal{N}_{\text{uncovered}} = \sum_{k=0}^{\mathcal{N}-1} \mathbb{I}(\Phi(\mathbf{n}_k) = 0)$ denotes the count of directions receiving absolutely zero coverage, with $\mathbb{I}(\cdot)$

being the indicator function. Finally, the comprehensive FOV coverage metric Γ is formulated as:

$$\Gamma = \xi \cdot \left(\frac{1}{N} \sum_{k=0}^{N-1} \Phi(\mathbf{n}_k) \right). \quad (29)$$

Here the $1/N$ normalization yields an average quality-weighted intensity, while the penalty multiplier $\xi \in [0, 1]$ scales down the score whenever blind spots exist, driving the optimization toward seamless spatial coverage.

Given the initial UAV set U_M from Alg. 1, we formulate the formation optimization problem to maximize FOV coverage while preserving perception performance. Crucially, we leverage the geometric symmetry of the observation manifold to perform formation transition. A central symmetry transition relocates a UAV to the antipodal position relative to the target, achieved by reversing all relative coordinates and shifting $\theta_i \rightarrow \theta_i + \pi$. Such a transition alters the signs of specific rows in the observation Jacobian $\mathbf{O}_i$, represented by a diagonal sign matrix $\mathbf{S}_i$ containing ± 1 on its diagonal, so that the transformed Jacobian is $\mathbf{O}'_i = \mathbf{S}_i \mathbf{O}_i$. Substituting the reversed coordinates and $\theta_i \rightarrow \theta_i + \pi$ into the camera Jacobian (Eq. (4)) yields $\mathbf{S}_{\text{cam}} = \text{diag}(-1, +1)$; likewise, for the LiDAR Jacobian (Eq. (13)), the same substitution yields $\mathbf{S}_{\text{lidar}} = \text{diag}(-1, -1, +1)$. The updated FIM becomes:

$$\mathcal{F}'_i = (\mathbf{S}_i \mathbf{O}_i)^T \mathbf{Q}^{-1} (\mathbf{S}_i \mathbf{O}_i) = \mathbf{O}_i^T (\mathbf{S}_i^T \mathbf{Q}^{-1} \mathbf{S}_i) \mathbf{O}_i. \quad (30)$$

The invariance of the aggregated FIM relies on the two independence assumptions established in Sec. III-B. First, the measurement covariance $\mathbf{Q}$ is strictly diagonal; since diagonal matrices commute, $\mathbf{S}_i^T \mathbf{Q}^{-1} \mathbf{S}_i = \mathbf{Q}^{-1}$ holds ($\mathbf{S}_i^T \mathbf{S}_i = \mathbf{I}$), preserving the per-UAV FIM ($\mathcal{F}'_i = \mathcal{F}_i$). Second, statistically independent cross-UAV noise ensures the aggregated FIM remains the linear sum of per-UAV FIMs, preserving global invariance [5]. Consequently, this central-symmetry reconfiguration is a mathematically exact operation that improves FOV coverage and communication topology without sacrificing the CRLB. If these conditions are violated (non-diagonal covariance, correlated noise, or large calibration errors), the exact cancellation degrades into a near-optimal heuristic. Hence, the optimization problem is:

$$\begin{aligned} \text{(P2):} \quad & \max_{\mathcal{P}_i \in \Omega_{\text{sym}}(U_M)} \Gamma \\ \text{s.t.} \quad & \eta_i^{\text{SINR}} \geq \eta_{\min}^{\text{SINR}}, \quad \forall i \in U_M \end{aligned}, \quad (31)$$

where $\eta_{\min}^{\text{SINR}}$ is the minimum required SINR ratio for reliable wireless communication. $\Omega_{\text{sym}}(U_M)$ is the discrete space of equivalent formations generated by applying the central symmetry transition to any subset of UAVs in U_M .

To address the combinatorial complexity of evaluating all 2^M possible flip allocations for M UAVs, we develop an efficient search strategy based on azimuth sector partitioning. The algorithm first computes the yaw angle θ_i of each UAV with respect to the target. Since each UAV is assumed to face the target, the yaw angle is determined by the bearing from the UAV to the target:

$$\theta_i = \text{atan2}(y_{\text{tar}} - y_i, x_{\text{tar}} - x_i), \quad (32)$$

where $\text{atan2}(y, x)$ denotes the two-argument arctangent function, which returns the angle of the vector (x, y) in the full range $(-\pi, \pi]$ and correctly handles all quadrants. These angles are then partitioned into K sectors according to:

$$\text{sector}_i = \lfloor \frac{\theta_i}{2\pi/K} \rfloor. \quad (33)$$

Within sectors containing multiple UAVs (typically exhibiting high coverage overlap), we prioritize flipping operations for UAVs that yield the greatest marginal improvement in Γ while satisfying the SINR constraint.

Alg. 2 details this formation optimization procedure, which systematically evaluates candidate flips while pruning the search space through sector-based constraints. This approach efficiently identifies the UAV formation that maximizes FOV coverage completeness without compromising perception accuracy or communication reliability.

Algorithm 2: FOV-oriented Formation Optimization

```

1 Input: Formation  $U$  from Alg. 1, target position  $\mathbf{p}_t$ 
2 Output: Optimal formation  $U^*$ 
3  $U^* \leftarrow U, \mathcal{T}^* \leftarrow \Gamma(U)$ 
4 for each  $u_i \in U$  do
5    $\theta_i \leftarrow \text{atan2}(y_{\text{tar}} - y_i, x_{\text{tar}} - x_i),$ 
6    $\text{sector}_i \leftarrow \lfloor \frac{\theta_i}{2\pi/K} \rfloor$ 
7 for each  $\text{sector}_k$  with multiple UAVs do
8   for each  $u_i \in \text{sector}_k$  do
9      $U' \leftarrow \text{Flip}(U, u_i),$ 
10     $\mathcal{T}' \leftarrow \Gamma(U')$ 
11    if  $\eta_i^{\text{SINR}}(U') \geq \eta_{\min}^{\text{SINR}}$  and  $\mathcal{T}' > \mathcal{T}^*$  then
12       $U^* \leftarrow U'$ 
13       $\mathcal{T}^* \leftarrow \mathcal{T}'$ 
```

C. Flight Control

Building upon the optimal FOV formation derived in Sec. IV-B, this subsection addresses the challenge of transitioning the UAV swarm from its initial position to the desired geometric formation $\mathcal{P}^* = [\mathcal{P}_1^*, \mathcal{P}_2^*, \dots, \mathcal{P}_m^*]^T$ while ensuring energy efficiency, trajectory smoothness, and asymptotic stability. Conventional control approaches often prioritize rapid convergence at the expense of energy consumption and maneuver aggressiveness. In contrast, we develop a Lyapunov-based distributed control framework that systematically balances formation accuracy with operational efficiency. The proposed control architecture incorporates a logarithmic potential field to bound control authority during large formation errors, thereby preventing excessive maneuvers while guaranteeing global asymptotic stability.

At the beginning, we model each UAV's dynamics using a double integrator system:

$$\dot{\mathcal{P}}_i(t) = \mathbf{v}_i(t), \quad (34)$$

$$m_i \dot{\mathbf{v}}_i(t) = \mathbf{u}_i(t), \quad (35)$$

where $\mathcal{P}_i$, $\mathbf{v}_i$, and m_i represent the position, velocity, and mass of the i -th UAV, respectively, and $\mathbf{u}_i(t)$ denotes the control input. Additionally, the desired formation is characterized by target relative displacements d_{ij} between neighboring UAV pairs (i, j) , satisfying $\mathcal{P}_i^* - \mathcal{P}_j^* = d_{ij}$ where $d_{ij} = -d_{ji}$. Then, the formation error between UAVs i and j is defined as:

$$\mathbf{e}_{ij}(t) = \mathcal{P}_i(t) - \mathcal{P}_j(t) - d_{ij}, \quad (36)$$

where $\mathcal{P}_i^*$ indicates the desired position of the i -th UAV. The formation objective requires $\mathbf{e}_{ij} \rightarrow 0$ for all connected pairs. The communication topology is denoted by a connected undirected graph $\mathcal{G}$, with $\mathcal{M}_i$, the neighbor set of UAV i .

Based on these dynamics, we design the control scheme in a leader-follower structure and then verify its stability via a Lyapunov argument. Specifically, a designated leader UAV L incorporates trajectory tracking capability, while the follower UAVs execute distributed formation control. The proposed control scheme is formulated as:

$$\mathbf{u}_i = \begin{cases} -k_1 \sum_{j \in \mathcal{M}_L} \frac{\mathbf{e}_{Lj}}{1 + \|\mathbf{e}_{Lj}\|^2} & i = L, \\ -k_2 \mathbf{v}_L - k_p(\mathcal{P}_L - \mathcal{P}_L^{\text{des}}), & \\ -k_1 \sum_{j \in \mathcal{M}_i} \frac{\mathbf{e}_{ij}}{1 + \|\mathbf{e}_{ij}\|^2} - k_2 \mathbf{v}_i, & i \neq L, \end{cases} \quad (37)$$

where $\mathcal{P}_L^{\text{des}}$ represents the leader's desired absolute position (determined by the tracked target), and $k_1, k_2, k_p > 0$ are positive control gains. The stability of the closed-loop system is established using Lyapunov theory. Consider the candidate Lyapunov function:

$$\mathcal{L} = \frac{k_1}{2} \sum_{(i,j) \in \mathcal{G}} \ln(1 + \|\mathbf{e}_{ij}\|^2) + \sum_i \frac{1}{2} \|\mathbf{v}_i\|^2 + \frac{k_p}{2} \|\mathcal{P}_L - \mathcal{P}_L^{\text{des}}\|^2. \quad (38)$$

This positive definite function combines three key components: (i) a logarithmic potential field $\ln(1 + \|\mathbf{e}_{ij}\|^2)$ that bounds control authority during large formation errors, preventing excessive maneuvers; (ii) a velocity damping term $\|\mathbf{v}_i\|^2$ that suppresses oscillations and ensures smooth velocity profiles; (iii) a trajectory tracking component $\frac{k_p}{2} \|\mathcal{P}_L - \mathcal{P}_L^{\text{des}}\|^2$ for the leader UAV L . Together, these elements synergistically balance convergence speed with energy efficiency while guaranteeing formation accuracy. Hereafter, differentiating Eq. (38) with respect to time yields:

$$\dot{\mathcal{L}} = k_1 \sum_{i=1}^M \left(\sum_{j \in \mathcal{M}_i} \frac{\mathbf{e}_{ij}}{1 + \|\mathbf{e}_{ij}\|^2} \right)^T \mathbf{v}_i + \sum_{i=1}^M \mathbf{v}_i^T \mathbf{u}_i + k_p(\mathcal{P}_L - \mathcal{P}_L^{\text{des}})^T \mathbf{v}_L. \quad (39)$$

Substituting the control law (Eq. (37)) into Eq. (39), the formation-error term and the leader-tracking term cancel identically for any $k_1, k_2, k_p > 0$, yielding:

$$\dot{\mathcal{L}} = -k_2 \sum_{i=1}^M \|\mathbf{v}_i\|^2 \leq 0. \quad (40)$$

By LaSalle's invariance principle [26], the system is globally asymptotically convergent, ensuring all UAVs converge to their designated positions in the desired formation. The logarithmic potential field in the Lyapunov function specifically optimizes trajectory smoothness by preventing control saturation during large formation errors, thereby minimizing energy-intensive maneuvers while maintaining precise formation tracking. This design achieves an optimal balance between formation accuracy, trajectory smoothness, and energy efficiency for UAV swarm operations.

V. NUMERICAL SIMULATION

This section numerically validates the proposed framework along three modules: the UAV-sensor allocation, the FOV-oriented optimal formation, and energy-efficient flight control. It quantifies its improvements over conventional benchmarks under practical constraints of communication interference, energy limitations, and hardware budgets.

All experiments were conducted on an Ubuntu 24.04 workstation with an Intel Core i7-13650HX CPU, 24 GB DDR5 RAM, and an NVIDIA RTX 4060 GPU, using Python 3.12. The task-level validation (Section V-D) used Gazebo 11.10 with ROS Noetic.

To ensure reproducibility, the critical simulation settings are listed in Tab. I. The key parameters governing the optimization, communication, and control layers are explicitly defined. The objective weighting coefficients (w_1, w_2) prioritize perception information gains during initial iterations while penalizing redundant hardware allocation. For the greedy selection algorithm, the termination threshold (ρ) is calibrated to balance perception accuracy with system overhead. Furthermore, the flight control gains (k_1, k_2, k_p) were optimized via grid search to guarantee smooth formation transitions.

TABLE I: Key Simulation Parameters

Parameter Description	Symbol	Value
Maximum candidate UAVs	$M_{\max}$	10
Target velocity vector	$\mathbf{v}_{\text{tar}}$	$[0.5, 0.3, 0.0]^T$ m/s
UAV mass	m_i	1.0 kg
FOV azimuth evaluation step	$\Delta\phi$	10° (36 directions)
Carrier frequency	f_c	2.4 GHz
Transmit power	P_t	20 dBm
Minimum SINR threshold	$\eta_{\min}^{\text{SINR}}$	10 dB
Path loss exponent	α	2.0
Noise power	σ^2	-110 dBm
Camera focal length	f_x, f_y	381.0 pixels
LiDAR measurement covariance	$\mathbf{Q}_{\text{lidar}}$	$\text{diag}(0.1, 0.02, 0.015)$
Camera measurement covariance	$\mathbf{Q}_{\text{cam}}$	$\text{diag}(6, 6)$
FOV limits	γ, κ	$50^\circ, 40^\circ$
Angular discretization step	$\Delta\beta, \Delta\delta$	$10^\circ, 10^\circ$
Sensor hardware cost	$C^{\text{cam}}, C^{\text{lidar}}$	0.7, 2.5
Azimuth sector count	K	8
Objective weighting coefficients	w_1, w_2	0.18, 0.20
Greedy termination threshold	ρ	0.17
Coverage metrics	$\sigma_\gamma, d_{\text{ref}}$	$25^\circ, 10$ m
Flight control gains	k_1, k_2, k_p	4.0, 1.5, 10.0

A. UAV-Sensor Allocation

We implemented the proposed greedy algorithm detailed in Sec. IV-A to determine the optimal allocation of UAVs

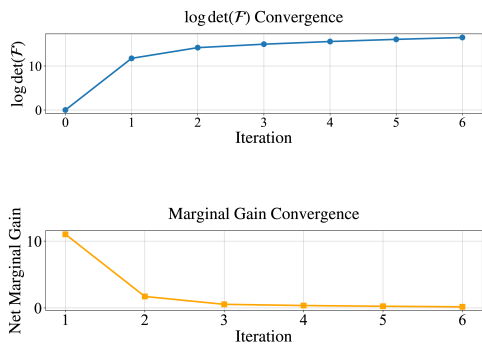


Fig. 4: Convergence of the greedy UAV-sensor allocation. Top: $\log \det(\mathcal{F})$ saturates as the number of UAVs increases. Bottom: the net marginal gain diminishes accordingly, confirming the submodularity of the objective.

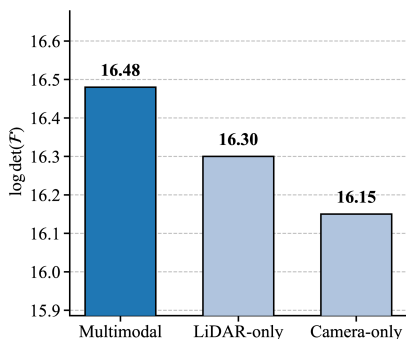


Fig. 5: Performance between multi- and single-modal formation.

and sensor modalities. The objective function maximizes $\log \det \mathcal{F}(C)$, while incorporating penalty terms for resource consumption $\mathcal{C}^{comm}$ and hardware budgets C^{sensor} .

Fig. 4 shows the convergence of the Alg. 1. As expected from the submodular property of the FIM determinant, the perception accuracy shows diminishing returns as the number of UAVs of the formation increases. This result justifies the termination criterion for the greedy algorithm (Alg. 1). Alg. 1 yields an optimal UAV-sensor allocation comprising four camera-equipped UAVs and two LiDAR-equipped UAVs. The corresponding formation geometry is visualized in Fig. 9a, with parameters provided in Tab. II.

Fig. 5 provides the performance comparison between multimodal and single-modal formation by applying Alg. 1. Here, the multimodal case allows both cameras and LiDARs, whereas the single-modal case restricts UAVs to those equipped only with cameras or LiDARs. The optimal multimodal formation attains a $\log \det(\mathcal{F})$ of 16.48 with only 6 UAVs (2 LiDARs, 4 cameras). However, the LiDAR-only and camera-only allocations reach 16.30 (4 LiDARs) and 16.15 (7 cameras). This shows that multimodal formation can improve perceptual accuracy through the sensor diversity.

Furthermore, we benchmark Alg. 1 against a Monte Carlo (MC) search and Particle Swarm Optimization (PSO). Here, PSO is initialized with 30 particles over 50 iterations, and the MC search is configured with 50 independent stochastic trials. The reported MC and PSO results are averaged over 30 independent runs to account for stochastic variability; the

proposed greedy algorithm is deterministic and thus requires a single execution. As illustrated in Fig. 6, our approach locates a highly precise solution ($\log \det(\mathcal{F}) = 16.48$) in just 2.3 seconds. Conversely, PSO prematurely converges to a suboptimal local maximum (≈ 16.44) after 3.5 seconds, and the MC search plateaus at a comparable sub-optimal bound even after 20.0 seconds. Evaluated on the pure FIM objective $f(U) = \log \det \mathcal{F}$, Alg. 1 satisfies the $(1 - 1/e)$ -approximation bound ($16.482 > 12.57$) compared to the global optimum (19.884) obtained via exhaustive search. Our proposed algorithm bypasses both combinatorial explosion and local optima traps, making it highly suitable for real-time SWaP-constrained UAV deployments.

To analyze the sensitivity to discretization, we conducted comparative experiments with varying angular resolutions. As shown in Fig. 7, the coarse discretization (e.g., $\Delta\beta = 60^\circ, \Delta\delta = 30^\circ$) fails to identify optimal allocations, resulting in degraded perception accuracy. The moderate discretization (e.g., $\Delta\beta = 10^\circ, \Delta\delta = 10^\circ$) achieves an optimal balance between computational efficiency and solution quality. Notably, elevation angle discretization shows greater influence on accuracy than azimuth discretization, providing practical guidance for real-world implementations.

In practical, minor timing misalignments and extrinsic calibration offsets are characterized as zero-mean additive perturbations that inflate the baseline measurement covariance matrices ($\mathbf{Q}_{cam}$ and $\mathbf{Q}_{lidar}$). To evaluate this impact, an equivalent noise inflation factor is applied to scale the covariance matrices from a 1.0x ideal baseline to a 2.0x extreme uncertainty condition. Test A (Fixed Baseline Config.) where the optimal formation is derived at the 1.0x baseline, and Test B (Re-optimized Config.) where the algorithm re-computes the optimal formation under the inflated noise. Fig. 8 gives the cumulative information gain that is quantified by $\log \det(\mathcal{F})$. This exhibits a predictable degradation (from 16.48 down to 12.32) as the equivalent noise intensifies.

Remarkably, the performance degradation curves from both tests coincide. The dynamic re-optimization in Test B confirms that the optimal sensor allocation remains strictly invariant (comprising 2 LiDARs and 4 cameras) across the entire noise spectrum. The performance equivalence between the static and re-optimized topologies indicates that the original optimal formation retains a robust information geometry within the evaluated noise range. Consequently, within the evaluated range of equivalent diagonal covariance inflation, the swarm maintains its optimal formation topology without necessitating immediate spatial reconfigurations.

TABLE II: UAV formation settings

UAV ID	Sensor	Azimuth ($^\circ$)	Elevation ($^\circ$)	Position (x, y, z)
1	LiDAR	40.0	160.0	(-7.2, -6.0, 3.4)
2	LiDAR	130.0	20.0	(-6.0, 7.2, 3.4)
3	Camera	0.0	160.0	(-9.4, -0.0, 3.4)
4	Camera	100.0	20.0	(-1.6, 9.3, 3.4)
5	Camera	50.0	160.0	(-6.0, -7.2, 3.4)
6	Camera	140.0	160.0	(7.2, -6.0, 3.4)

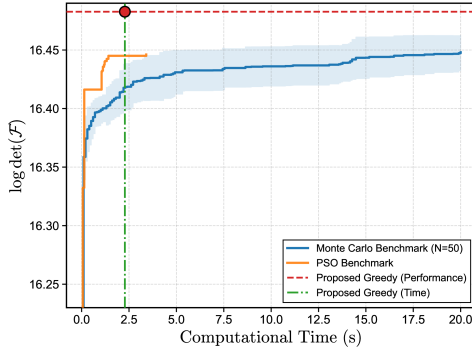


Fig. 6: Computational Efficiency Comparison: Proposed Greedy Algorithm vs. Monte Carlo and PSO Benchmarks

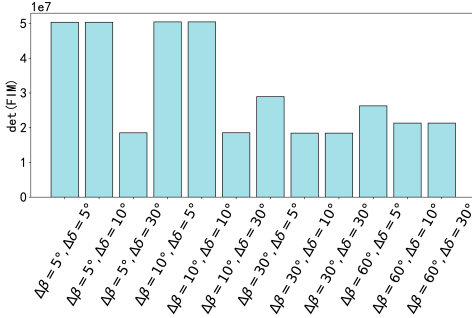


Fig. 7: Effect of Angular Discretization on Perception Performance

B. FOV-oriented Formation Performance

Upon the UAV-sensor allocation derived in Sec. V-A, Alg. 2 employs sector-constrained flipping operations (with $K = 8$ azimuthal sectors) to explore the equivalent formation space while maintaining computational tractability. Fig. 9b shows the optimal formation for aerial target tracking. Alg. 2 successfully transforms the initial UAV-sensor allocation into a FOV-oriented formation that eliminates coverage blind spots while preserving the FIM determinant. This spatial reconfiguration shows how equivalent geometric transitions can improve task effectiveness without compromising theoretical bounds.

For ground SAR scenarios, terrain constraints preclude UAV deployment below the target plane. To address this practical limitation, we implement a Z-axis reflection transition that repositions all UAVs to the feasible half-space while main-

taining formation integrity. Although this constrained transition introduces a marginal reduction in sensing accuracy, the resulting allocation (Fig. 9c) achieves significantly enhanced FOV coverage and communication reliability compared to the initial allocation.

Moreover, Tab. III provides a performance comparison between our proposed formation and a benchmark uniform enclosure allocation. Since Alg. 2 reconfigures the initial UAV-sensor allocation from Alg. 1 through the equivalent transition, the FOV and SINR gains reported against Alg. 1 reflect the effect of this transition alone, with the FIM determinant held fixed. The uniform enclosure (Bch.) places the same 6 UAVs (two LiDARs and four cameras) evenly around the target as an external geometric baseline, so that every reported percentage stems from formation geometry rather than resource endowment. For aerial targets, the optimal formation (Alg. 2) maintains the FIM log-determinant (Alg. 1) at 16.48 (superior to the uniform enclosure's 16.124), while simultaneously achieving a 25.0% improvement in FOV coverage metric Γ (increasing from 0.61 (Alg. 1) to 0.76 (Alg. 2)). Most significantly, the average SINR increases by 104.2% (from 6.97 (Alg. 1) to 14.23 dB (Alg. 2)), with the minimum SINR rising from 1.78 dB (Alg. 1) to 11.51 dB (Alg. 2), thereby ensuring reliable communication across the entire formation control.

In ground target scenarios, the Z-axis reflection transition yields a log-determinant of 16.414, marginally lower than the aerial target allocation but still exceeding the uniform enclosure's 16.241. The FOV coverage metric reaches 0.76, surpassing the uniform enclosure's 0.74, while the average SINR improves by 97.3% relative to the initial state (reaching 13.78 dB). Although slightly lower than the uniform enclosure's 14.54 dB, this SINR level remains well within operational requirements for reliable data transmission.

TABLE III: Performance of different formations

Metric	Alg1	Alg2 (Air)	Alg2 (Gnd)	Bch. (Gnd)	Bch. (Air)
$\log \det(\mathcal{F})$	16.48	16.48	16.41	16.24	16.12
Γ	0.61	0.76	0.76	0.74	0.74
Avg. SINR	6.97	14.23	13.78	14.54	14.92
Min. SINR	1.78	11.51	11.02	12.46	12.89

Note: Alg1 is the formation from Alg. 1; Alg2 is the formation from Alg. 2 (Air/Ground); Bch.: Uniform enclosure formation (Ground/Air).

These results show that formation geometry optimization can simultaneously enhance multiple performance metrics, i.e., coverage completeness, communication reliability, and preservation of sensing accuracy. This supports our hypothesis that 3D spatial formation is an important rather than incidental component in multimodal swarm perception systems.

C. Flight Control Efficiency

To validate the theoretical stability and energy efficiency of the proposed Lyapunov control in Sec. IV-C, we conducted simulations involving three distinct control strategies: (1) the proposed Lyapunov control with logarithmic potential function, (2) the conventional Lyapunov control with quadratic potential function, and (3) the classical Artificial Potential Field (APF) method. The simulation scenario includes a formation

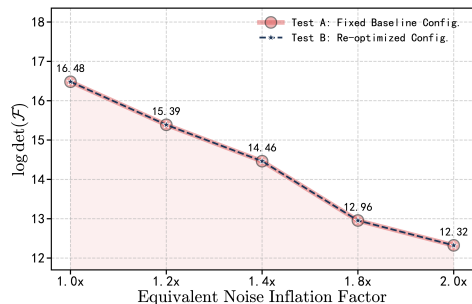


Fig. 8: Dual-track sensitivity of $\log \det(\mathcal{F})$ under equivalent noise inflation from $1.0\times$ to $2.0\times$, for the fixed-baseline (Test A) and re-optimized (Test B) configurations.

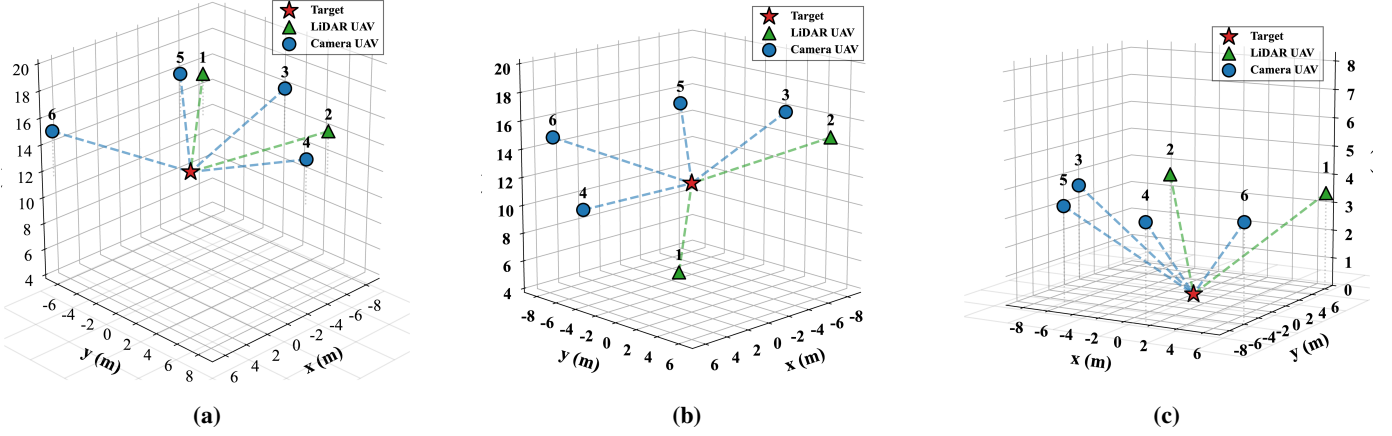


Fig. 9: Heterogeneous UAV formation allocations (four camera-equipped UAVs as blue circles, two LiDAR-equipped UAVs as green triangles, target as red star): (a) initial formation produced by Alg. 1; (b) formation for an aerial target; and (c) formation for a ground target.

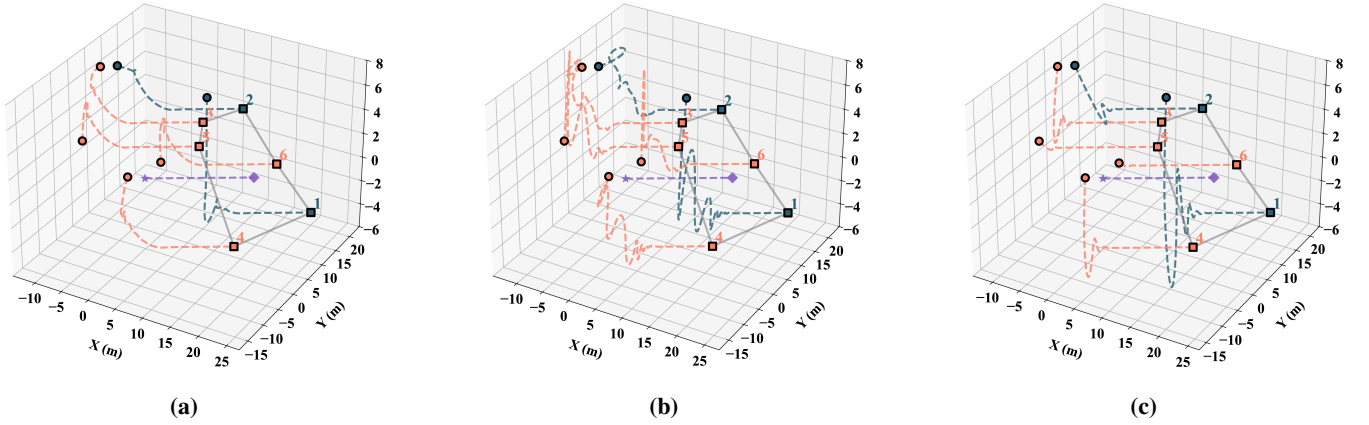


Fig. 10: 3D trajectory evolution of the six UAVs during formation transition under (a) the proposed logarithmic-potential controller, (b) the quadratic-potential controller, and (c) the classical APF method. Purple star, dashed line, and diamond represent the target initial position, trajectory, and final position, respectively. Circles denote the UAV initial positions. Squares (labeled 1–6, matching Tab. II) denote the final positions. Gray lines represent the target formation. Blue and orange point represent the LiDAR UAVs and camera UAVs, respectively.

comprising 6 UAVs ($M = 6$) with the optimal allocation derived from Alg. 2. All three controllers are deterministic under fixed initial conditions; the reported metrics correspond to a single execution per method. To emulate a dynamic tracking of SAR operations, the formation tracks a maneuvering target trajectory $\mathcal{P}_L^{des}(t)$ with continuous velocity and directional variations. Moreover, all three flight control schemes share the same target formation and the same target reference trajectory, so that the reported differences stem solely from the control scheme. Here, the quadratic control scheme is:

$$\mathbf{u}_i = \begin{cases} -k_1 \sum_{j \in \mathcal{M}_i} (\mathbf{e}_{ij}) - k_2 \mathbf{v}_i - k_p (\mathcal{P}_i - \mathcal{P}_i^{des}), & i = L, \\ -k_1 \sum_{j \in \mathcal{M}_i} (\mathbf{e}_{ij}) - k_2 \mathbf{v}_i, & i \neq L, \end{cases} \quad (41)$$

where L is the indicator of the leader UAV. In contrast, the APF method applies an attractive force guiding the UAV toward the desired position and a repulsive force for collision

avoidance. The total APF force input is:

$$\mathbf{u}_i = -k_a (\mathcal{P}_i - \mathcal{P}_i^*) + \sum_{j \in \mathcal{M}_i} k_r \left(\frac{1}{\|\mathcal{P}_{ij}\|} - \frac{1}{d_0} \right) \frac{\mathcal{P}_{ij}}{\|\mathcal{P}_{ij}\|^3} - k_2 \mathbf{v}_i, \quad (42)$$

where $\mathcal{P}_{ij} = \mathcal{P}_i - \mathcal{P}_j$ and d_0 is the distance of safe separation. The middle repulsive term is only active when $\|\mathcal{P}_{ij}\| < d_0$ and evaluates to zero when $\|\mathcal{P}_{ij}\| \geq d_0$. Furthermore, k_a and k_r are the attractive and repulsive gains. k_2 is the velocity damping gain to suppress oscillations.

TABLE IV: Performance of different flight control

Metric	Logarithmic	Quadratic	Classical APF
Avg. Distance (m)	15.62	29.60	19.40
Avg. Velocity Err.	0.678	1.736	0.883
Max. Velocity Err.	7.666	8.907	10.865
Avg. Final Pos. Err. (m)	0.056	0.085	0.107
E_{total} (m ² /s ³)	437.7	2858.8	1043.5

Tab. IV summarizes the performance metrics during the formation transition. To evaluate control efficiency, the total

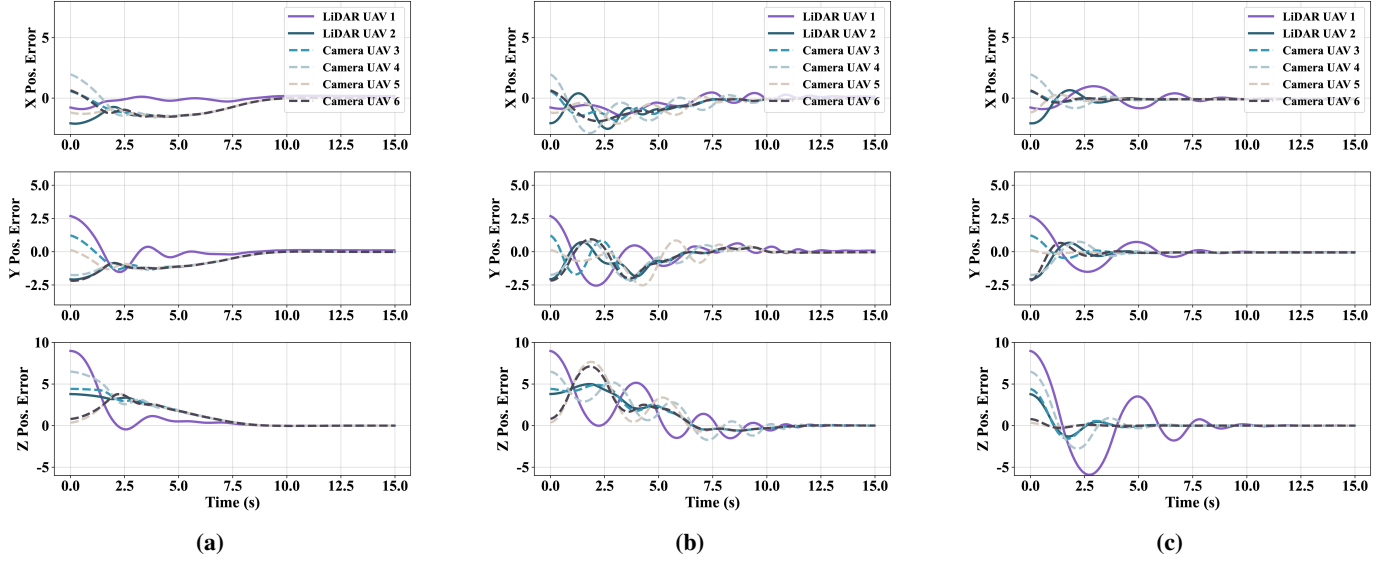


Fig. 11: 3D position error evolution (X , Y , Z) of the six UAVs during formation transition under (a) the proposed logarithmic-potential controller; (b) the quadratic-potential controller; and (c) the classical APF method. Solid lines: LiDAR UAVs; dashed lines: camera UAVs.

control effort E_{total} over the transition period T is defined as:

$$E_{\text{total}} = \sum_{i=1}^M \int_0^T \|\mathbf{u}_i(t)\|^2 dt, \quad (43)$$

where E_{total} measures the integral-square control input. Because actuator power scales monotonically with $\|\mathbf{u}_i(t)\|^2$, a reduction in E_{total} directly translates to lower propulsion energy consumption, though the proportionality constant depends on platform-specific aerodynamic and motor efficiency parameters. Minimizing E_{total} involves a dual-factor optimization: suppressing the instantaneous control magnitude $\|\mathbf{u}_i(t)\|^2$ and reducing the total time T . The proposed logarithmic control scheme achieves both: its bounded control authority yields a smooth, near-constant velocity profile that restricts $\|\mathbf{u}_i(t)\|^2$, while the shortest travel trajectory proportionally reduces T .

The empirical results substantiate this dual-factor efficiency. Regarding T , the proposed method achieves the shortest average travel distance of 15.62 m, a 47.2% reduction over the quadratic controller (29.60 m) and 19.5% over the classical APF (19.40 m). Regarding the control magnitude, it attains an average velocity error of 0.678 m/s and a maximum of 7.666 m/s, lower than the quadratic (1.736 and 8.907 m/s) and APF (0.883 and 10.865 m/s) methods, alongside the smallest final position error of 0.056 m. Combining both factors, the control effort E_{total} evaluated from the recorded inputs is only 437.7 m²/s³ for the proposed method—an 84.7% and 58.1% reduction over the quadratic and APF baselines. It confirms that the logarithmic control scheme lowers control effort and, consequently, propulsion energy consumption, not only by shortening the trajectory but also by bounding the instantaneous control magnitude $\|\mathbf{u}_i(t)\|^2$.

Fig. 10 shows the 3D trajectory evolution with various control schemes. The proposed logarithmic scheme (Fig. 10a) exhibits smooth convergence without oscillatory behavior. In contrast, both the quadratic (Fig. 10b) and APF scheme (Fig. 10c) have pronounced oscillations during the transient

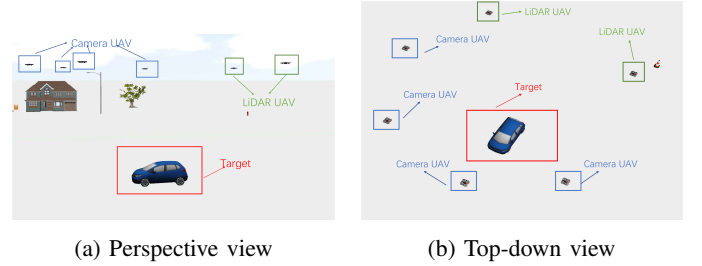


Fig. 12: Gazebo settings: (a) perspective view; (b) top-down view of the optimal formation (two LiDARs and four cameras).

TABLE V: Runtime of three modules

Module	Setting	Runtime
UAV-sensor allocation	$ V = 1368$	2.30 s
Formation optimization	$M = 6$	1.13 s
Flight-control update	$M = 6$, per step	1.11 ms

phase. These oscillations necessitate additional corrective maneuvers that increase energy consumption and flight instability.

The position error dynamics in Fig. 11 provide further evidence of the proposed (logarithmic) control scheme's advantages. The proposed scheme achieves monotonic error reduction across all three spatial dimensions with minimal fluctuations, whereas both comparative methods exhibit significant oscillatory behavior before reaching steady state. While maintaining comparable convergence rates to the quadratic scheme, our approach eliminates the high-frequency oscillations that characterize conventional methods during the transient phase. This smooth error evolution directly contributes to reduced control effort, as evidenced by the lower maximum velocity error, and minimizes energy-intensive maneuvers.

Finally, we assess the real-time feasibility of the three modules. As reported in Tab. V, the one-shot allocation and formation-optimization steps complete within a few seconds,

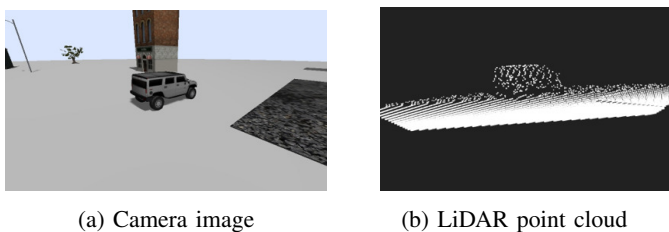


Fig. 13: Example raw sensor data in Gazebo: (a) RGB image from a camera-equipped UAV; (b) Point cloud from a LiDAR-equipped UAV.

while the continuously running flight-control update takes only 1.11 ms per step, far exceeding the 50–100 Hz rate of practical controllers. Since the flight-control scheme requires only lightweight vector operations with a small memory footprint, it suits resource-constrained embedded controllers, whereas the one-shot steps can be executed pre-flight or on edge servers.

D. Task-Level Validation via Gazebo

While the preceding evaluations validate the proposed framework through the information-optimal metric $\log \det(\mathcal{F})$, we further conduct a high-fidelity, task-level validation in the Gazebo simulation environment to bridge the gap between the geometric optimality and the cooperative perception performance. Typically, Gazebo is a widely adopted robotics simulator that provides a high-fidelity physics engine and photorealistic sensor rendering, enabling realistic emulation of UAV dynamics, camera imaging, and LiDAR point-cloud acquisition. It thus offers a faithful testbed for evaluating multi-UAV cooperative perception prior to costly physical deployment. To this end, we deploy the information-optimal formation derived from our framework into a photorealistic Gazebo world, where the 6 UAVs cooperatively perceive a ground target. Fig. 12 illustrates the constructed scenario from two complementary viewpoints: a perspective view depicting the realistic sensing environment (Fig. 12(a)), and a top-down view showing the optimal formation geometry around the target (Fig. 12(b)). Fig. 13 shows raw observations captured by a camera-equipped and a LiDAR-equipped UAV, respectively. The multi-view fusion is performed using a distributed multi-view learning backbone [27], where each UAV provides a single modality according to its onboard sensor: the two LiDAR UAVs encode their point clouds with PointNet, while the four camera UAVs encode their images with DenseNet121. The resulting 6 modality embeddings are then aggregated by a graph convolutional integrator to produce the final output. The perception task is target recognition: given the fused multimodal observations, the formation identifies the target.

To reflect adverse real-world conditions, we inject additive Gaussian noise into the raw images and point clouds produced by the Gazebo sensors, thereby corrupting the inputs to the cooperative perception. Note that this data-level perturbation is distinct from the measurement covariance $\mathbf{Q}$ in Tab. I: the latter characterizes the per-measurement statistical precision used in the FIM-based formation selection, whereas the former is applied to the raw sensor data to stress-test the end-to-end perception pipeline.

TABLE VI: Perception accuracy in Gazebo.

	Image Noise		Point-Cloud Noise		Hemispherical	Optimal
	mean	std	mean	std		
S1	0.04	0.02	0.02	0.05	82.0%	94.4%
S2	0.06	0.02	0.04	0.05	61.0%	73.3%

Hereafter, we evaluate the system under two test scenarios with progressively higher noise intensities. In Scenario 1 (S1), the image noise has mean 0.04 and standard deviation 0.02, while the point cloud noise has mean 0.02 and standard deviation 0.05. In scenario 2 (S2), both noise sources are further intensified, with the image noise mean raised to 0.06 and the point-cloud noise mean raised to 0.04, while the standard deviations remain at 0.02 and 0.05, respectively.

Under these settings, we benchmark the perception accuracy of the proposed optimal formation against 100 hemispherically sampled formations drawn from the same discrete candidate set V defined in Sec. IV-A, each consisting of 6 UAVs drawn under the identical heterogeneous sensor budget (two LiDARs and four cameras). For a fair comparison, each formation is evaluated over 10 independent noise realizations per scenario, with all formations sharing the same set of noise seeds so that any performance difference is attributable solely to formation geometry. As summarized in Tab. VI, Scenario 1 shows that the optimal formation attains an accuracy of 94.4%, surpassing the average of the hemispherical formations (82.0%) by 12.4 percentage points (a relative improvement of approximately 15.2%). In Scenario 2, the optimal formation still reaches 73.3%, outperforming the hemispherical formations (61.0%) by 12.3 percentage points (a relative improvement of approximately 20.2%). These task-level results confirm that the information-optimal geometry captures more discriminative multimodal features, yielding higher perception accuracy under noisy sensing conditions, with the benefit amplifying as observation quality deteriorates.

VI. CONCLUSION

This paper has shown that 3D formation geometry is an important rather than incidental design variable in multimodal UAV swarm cooperative perception during the active target tracking phase. By developing an information-optimal framework grounded in FIM determinant maximization, we jointly optimized heterogeneous sensor allocation, spatial allocation, and flight trajectory control. The framework integrates a submodular greedy algorithm that empirically satisfies the $(1 - 1/e)$ -approximation bound on the FIM objective, an equivalent formation transition that enhances FOV coverage by 25.0% and communication SINR by 104.2% without sacrificing sensing accuracy, and a Lyapunov-stable logarithmic-potential controller that reduces control effort by 84.7% over a quadratic-potential baseline. Upon noise-injection scenarios, the proposed formation improves perception accuracy by approximately 15.2% and 20.2% over 100 hemispherical formations from the same candidate set. While these evaluations provide supporting evidence for the framework, high-

fidelity simulation is not a substitute for physical deployment: real flight turbulence, on-board calibration, battery limits, and outdoor communication instability remain unmodeled. To facilitate reproducibility, the algorithm implementation is publicly available at <https://github.com/wuxingyu1231z/FIM-UAV-Multimodal-Formation-Optimization>. Future research will focus on hardware-in-the-loop flight experiments and adaptive formation reconfiguration under dynamic environmental occlusions.

REFERENCES

- [1] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on uavs for wireless networks: Applications, challenges, and open problems," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2334–2360, 2019.
- [2] S. Lee, D. Har, and D. Kum, "Drone-assisted disaster management: Finding victims via infrared camera and lidar sensor fusion," in *2016 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, 2016, pp. 84–89.
- [3] N. Souli, P. Kardaras, P. Kolios, and G. Ellinas, "A joint rogue drone detection and tracking fusing doa and passive radar measurements," in *2024 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2024, pp. 1200–1207.
- [4] R. Tallamraju, E. Price, R. Ludwig, K. Karlapalem, H. H. Bühlhoff, M. J. Black, and A. Ahmad, "Active perception based formation control for multiple aerial vehicles," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4491–4498, 2019.
- [5] S. Zhao, B. M. Chen, and T. H. Lee, "Optimal Sensor Placement for Target Localization and Tracking in 2D and 3D," *International Journal of Control*, vol. 86, no. 10, pp. 1687–1704, Oct. 2013.
- [6] A. O. Hero and D. Cochran, "Sensor management: Past, present, and future," *IEEE Sensors Journal*, vol. 11, no. 12, pp. 3064–3075, 2011.
- [7] J. Wu, D. Li, Y. Yu, L. Gao, J. Wu, and G. Han, "An attention mechanism and adaptive accuracy triple-dependent maddpg formation control method for hybrid uavs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 9, pp. 11 648–11 663, 2024.
- [8] T. Gavin, S. Lacroix, and M. Bronz, "Multi-agent reinforcement learning based drone guidance for n-view triangulation," in *2024 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2024, pp. 578–585.
- [9] D. Liu, B. Fei, W. Bao, X. Zhu, and X. Li, "Dawn: Dynamic task planning of multi-uav with two-layer optimization mechanism in uncertain environments," *IEEE Internet of Things Journal*, vol. 11, no. 23, pp. 37 813–37 830, 2024.
- [10] Q. Zhang, Y. Luo, H. Jiang, and K. Zhang, "Aerial edge computing: A survey," *IEEE Internet of Things Journal*, vol. 10, no. 16, pp. 14 357–14 374, 2023.
- [11] W. Lee, H. Bang, and H. Leeghim, "Cooperative localization between small uavs using a combination of heterogeneous sensors," *Aerospace Science and Technology*, vol. 27, no. 1, pp. 105–111, 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1270963812001101>
- [12] A. Kurt, N. Saputro, K. Akkaya, and A. S. Uluagac, "Distributed connectivity maintenance in swarm of drones during post-disaster transportation applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 9, pp. 6061–6073, 2021.
- [13] K. Panwar, G. Fatima, and P. Babu, "Optimal Sensor Placement for Hybrid Source Localization Using Fused TOA-RSS-AOA Measurements," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1–15, 2022.
- [14] Y.-C. Liu, J. Tian, N. Glaser, and Z. Kira, "When2com: Multi-Agent Perception via Communication Graph Grouping," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, Jun. 2020, pp. 4105–4114.
- [15] R. Xu, H. Xiang, Z. Tu, X. Xia, M.-H. Yang, and J. Ma, "V2X-ViT: Vehicle-to-Everything Cooperative Perception with Vision Transformer," Aug. 2022.
- [16] V. Semenyuk, I. Kurmashev, A. Lupidi, D. Alyoshin, L. Kurmasheva, and A. Cantelli-Forti, "Advances in UAV detection: Integrating multi-sensor systems and AI for enhanced accuracy and efficiency," *International Journal of Critical Infrastructure Protection*, vol. 49, p. 100744, Jul. 2025.
- [17] P. Tian, P. Cheng, Y. Wang, Z. Wang, Z. Wang, M. Yan, X. Yang, and X. Sun, "UCDNet: Multi-UAV Collaborative 3D Object Detection Network by Reliable Feature Mapping," Jun. 2024.
- [18] D. Qiao and F. Zulkernine, "CoBEVFusion: Cooperative Perception with LiDAR-Camera Bird's-Eye View Fusion," Oct. 2023.
- [19] T. Feng, X. Wang, F. Han, L. Zhang, and W. Zhu, "U2UData: A Large-scale Cooperative Perception Dataset for Swarm UAVs Autonomous Flight," in *Proceedings of the 32nd ACM International Conference on Multimedia*, Oct. 2024, pp. 7600–7608.
- [20] H.-A. Hung, H.-H. Hsu, and T.-H. Cheng, "Optimal Sensing for Tracking Task by Heterogeneous Multi-UAV Systems," *IEEE Transactions on Control Systems Technology*, vol. 32, no. 1, pp. 282–289, Jan. 2024.
- [21] N. Sahu, L. Wu, P. Babu, B. S. M. R., and B. Ottersten, "Optimal Sensor Placement for Source Localization: A Unified ADMM Approach," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 4, pp. 4359–4372, Apr. 2022.
- [22] Z. Wang, R. Liu, Q. Liu, J. S. Thompson, and M. Kadoch, "Energy-efficient data collection and device positioning in uav-assisted iot," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1122–1139, 2020.
- [23] K. Nakai, K. Yamada, T. Nagata, Y. Saito, and T. Nonomura, "Effect of Objective Function on Data-Driven Greedy Sparse Sensor Optimization," *IEEE Access*, vol. 9, pp. 46 731–46 743, 2021.
- [24] F. Morbidi and G. L. Mariottini, "Active target tracking and cooperative localization for teams of aerial vehicles," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 5, pp. 1694–1707, 2013.
- [25] K. Yamada, Y. Saito, K. Nankai, T. Nonomura, K. Asai, and D. Tsubakino, "Fast Greedy Optimization of Sensor Selection in Measurement with Correlated Noise," vol. 158, p. 107619. [Online]. Available: <http://arxiv.org/abs/1912.01776>
- [26] J. LaSalle, "Some extensions of liapunov's second method," *IRE Transactions on circuit theory*, vol. 7, no. 4, pp. 520–527, 1960.
- [27] K. Xiong, R. Wang, S. Leng, Q. Zhao, and M. Peng, "Graph convolutional integration based distributed multi-view learning in urban air mobility," in *2023 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, 2023, pp. 1–6.



Kai Xiong received his Ph.D. degree from the University of Electronic Science and Technology of China (UESTC) in 2021. He is currently an Associate Professor with UESTC. His research interests include the design and optimization of next-generation intelligent mobile networks and urban air mobility/advanced air mobility.

Prof. Xiong is the recipient of the 2024 IEEE Leonard G. Abraham Prize. He serves as a Session Chair and a TPC Member for many international conferences.



Xingyu Wu received the B.S. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2025, where he is currently pursuing the M.S. degree in Information and Communication Engineering. His research interests include unmanned aerial vehicle swarm coordination, formation control in structured airspace, and multi-modal information fusion for cooperative sensing.



Anna Duan received the M.S. degree from Beihang University. She is an engineer of the AVIC Chengdu Aircraft Design and Research Institute. Her research interests include autonomous control of unmanned aerial vehicles and optimization of urban air mobility/advanced air mobility.



Yongjun Huang (Member, IEEE) received the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in June 2010 and December 2016, respectively.

From September 2013 to September 2015, he was a Visiting Scholar with the Solid-State Science and Engineering, and Mechanical Engineering, Columbia University, New York City, NY, USA, and a Visiting Project Scientist with the Electrical Engineering, University of California at Los Angeles (UCLA), Los Angeles, CA, USA.

He is currently an Associate Professor with the School of Information and Communication Engineering, UESTC. He has published more than 100 journal articles, more than 100 conference proceeding papers, eight invited talks, and one book chapter, with totally more than 3500 citations (H-index: 31, Google Scholar). His research interests include IoT hardware components (antennas, microwave passive components, and electromagnetic metamaterials) and chip-scale photonic crystal cavity optomechanics, low phase noise RF sources, and high-resolution force/field sensors.



Supeng Leng received the Ph.D. degree from Nanyang Technological University (NTU), Singapore, in 2005.

He is a Full Professor with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China. He is also the Director of the Sichuan International Joint Research Center for Ubiquitous Wireless Networks. He has been a Research Fellow with the Network Technology Research Center, NTU. He published over 200 research

papers and four books/book chapters in recent years. His research focuses on resource, spectrum, energy, routing, and networking in the Internet of Things, vehicular networks, broadband wireless access networks, and next-generation intelligent mobile networks.

Prof. Leng got the Best Paper Awards at four IEEE international conferences. He is the Editorial Member of four international journals and a reviewer for over 20 well-known academic international journals.



Jianhua He (Senior Member, IEEE) received the PhD degree from Nanyang Technological University, Singapore. He is a Professor with the University of Essex, UK.

His main research interests include 5G/6G wireless communications and networks, UAV, semantic communication, connected vehicles, autonomous driving, agentic AI and large language models, emergency response, Internet of Things, edge intelligence, intelligent transport systems, data analytics.

He published more than 200 research papers in international journals and conferences. He was the coordinator of EU projects COSAFE, VESAFE and SECOM on cooperative connected autonomous vehicles. He was a general chair of International conference on Autonomous Driving and Intelligent Interaction ICoADI'26, workshop chair of MobiArch'20 and ICAV'21. He was a member of the editorial board of IEEE Wireless Communications Letters and The Computer Journal.