2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

All-onboard Relative Positioning and Control Framework for Autonomous Micro-UAV Swarms Based on Vision-Optoelectronic-UWB Fusion and Distributed Graph Optimization

Chengsong Xiong[1], Jiaqi Wan[1], Qifan Tong[1], Wenshuai Lu[1], Qingning He[1], Zheng You[1]

Abstract— **The autonomous cooperation of micro unmanned aerial vehicle (UAV) swarms remains a key challenge. Existing swarm relative positioning and control methods demand high sensing, computing, and communication resources and rely on external equipment like GPS and ground stations. To address these issues, this paper proposes an all-onboard and external-aiding-free swarm relative measurement, positioning and control framework. The framework utilizes an onboard Vision-Optoelectronic-Ultra-Wideband (UWB) coupled measurement system to acquire inter-UAV relative distance and direction. Subsequently, the swarm's relative positions are solved via a distributed graph optimization (DGO) approach. Based on the solved relative positions, swarm cooperative control is implemented through a distributed Voronoi diagram approach. Experimental results demonstrate that the proposed method enables 150 g micro-UAVs to achieve nearly 100-meter autonomous outdoor formation flight and collaborative tracking of dynamic targets, with swarm relative localization accuracy reaching approximately 0.262 m. This work pioneers fully autonomous measurement and control for 100-gram scale UAV swarms without external infrastructure, significantly advancing autonomy and enabling swarm intelligence emergence.**

I. Introduction

To overcome the inherent resource constraints of micro unmanned aerial vehicles (UAVs), researchers form swarms to enable complex collaborative tasks [1], thus demonstrating capabilities far beyond those of individual units. Current research features few micro-UAV swarm systems capable of autonomous flight and task execution without relying on external infrastructure. The fundamental reason lies in the absence of lightweight technologies for inter-agent relative positioning and control.

Relative positioning provides essential coordinates for swarm cooperation [2]. To avoid the dependency of positioning methods on external equipment like GPS [3] and wireless base stations [4], inter-UAV relative positioning methods based on onboard sensors have gained increasing research interest, which mainly include: 1) pose estimation, featuring technologies such as visual-inertial odometry (VIO) [5] and optical flow [6]; 2) distance measurement, including Ultra-Wideband (UWB) technology [7] and laser time-of-flight (ToF) sensors [8]; 3) direction measurement, exemplified by monocular vision [9], optoelectronic [10], and beamforming [11]. While pose estimation methods fuse vision and IMU data to estimate UAV displacement, integration errors accumulate over time, causing significant drift. Therefore, state estimation results are usually fused with relative measurements among UAVs through Kalman filtering-based methods to improve accuracy [12], however, limitations in the field of view (FOV) of onboard sensors can lead to missing measurements in real-world scenarios. To address this issue, graph optimization (GO)-based methods that achieve omni-directional relative positioning are proposed [13]. In this scheme, the state of each robot node is represented as a vertex of the graph, and the edges of the graph represent the relative measurements between robots. Based on prior knowledge and the constraints provided by relative measurements, the positions of the robots can be solved through information exchange and optimization. However, the general GO method typically requires high-end UAVs with substantial computational power and high energy consumption.

Building upon relative positioning, a lightweight and efficient control method is also essential for swarm intelligence. Existing popular methods include consensus control [14] and distributed model predictive control [15], among others. However, their performance become suboptimal when applied to high-dynamic applications like cooperative target tracking. Such scenarios necessitate UAVs to conduct real-time analysis of both collaborative agents and target dynamics, rather than merely maintaining predefined formations. In recent years, methods based on Voronoi diagrams for swarm cooperative target tracking have been developed [16]. While promising, their real-world use is limited due to the lack of reliable onboard localization between UAVs and high computational cost.

Current research addresses key UAV swarm collaboration issues in isolation, lacking a unified lightweight framework for closed-loop autonomous control. This gap prevents satellite/base-station-independent operation. Moreover, centralized relative positioning architectures impose GPU-level computational demands, which are incompatible with general UAV swarms, especially gram-scale micro-UAVs.

In this study, we present an all-onboard, lightweight, and distributed swarm relative sensing and control framework based on multimodal sensing, distributed graph optimization (DGO) and Voronoi diagrams for real-world 100-gram scale micro-UAV swarms. The proposed framework uses onboard micro monocular cameras, optoelectronic detectors, UWB sensors, IMUs, and optical flow sensors for inter-

*This work was supported by the National Natural Science Foundation of China (Grant No.U21A6003).

[1]Authors are with the Department of Precision Instrument, Tsinghua University, Beijing, China. yz-dpi@mail.tsinghua.edu.cn, luwenshuai_dky@163.com

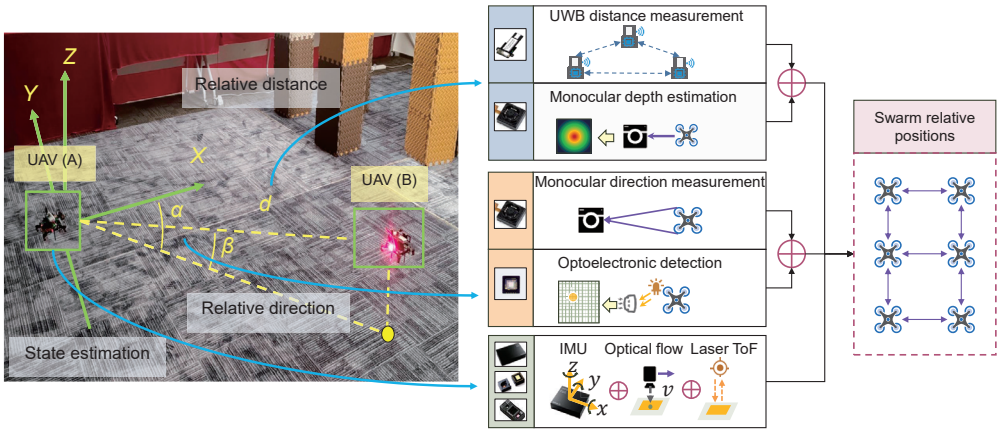

Fig. 1: Diagram of the multimodal onboard relative measurement system.

UAV relative measurements. Utilizing the sensing information and low-bandwidth inter-UAV communication, the relative positioning framework based on DGO runs on the lightweight ARM embedded computing units without relying on any external infrastructure. Integrating a lightweight object detection neural network, we implemented a swarm cooperative autonomous target tracking application. In outdoor field experiments, the framework achieved a relative swarm positioning accuracy of 0.262 m and a target localization accuracy of 0.43 m. The swarm also successfully demonstrated autonomous coordinated target encirclement through the proposed control method based on Voronoi graph. The integrated core sensing and control unit on each UAV consumes a total power of only 2.3 W and doesn't require a dedicated vision-processing GPU. This framework provides a viable solution for fully autonomous micro-UAV swarms.

II. Relative Positioning via Multimodal Sensing Fusion and Distributed Graph Optimization

In this section, we propose a lightweight Vision-Optoelectronic-UWB sensing system fusing relative distance and direction estimations. Utilizing the relative measurements, The swarm's relative positions are solved based on a DGO framework, exhibiting characteristics of omnidirectionality, global consistency, and strong real-time performance.

A. Multimodal Onboard Relative Sensing System

As illustrated in Fig. 1, the problem of 3D relative position measurement between UAVs is decomposed into the fusion of relative distance and direction measurements, assisted by an onboard state estimation system.

1) UWB Ranging: A distributed relative distance measurement network was achieved by mounting a UWB node (NOOPLOOP LinkTrack P-BS) on each UAV, without ground UWB anchors. The accuracy of UWB distance measurement can be influenced by various factors, including distance, signal power, and environmental electromagnetic interference. Therefore, we conducted a calibration process using the ground truth provided by a NOKOV Mars2HW optical motion capture system as reference. A linear transformation was applied to the raw output of UWB node to approximate the real distance. After calibration, the UWB operates at a frequency of 50 Hz, achieving a distance measurement accuracy within 0.1 m.

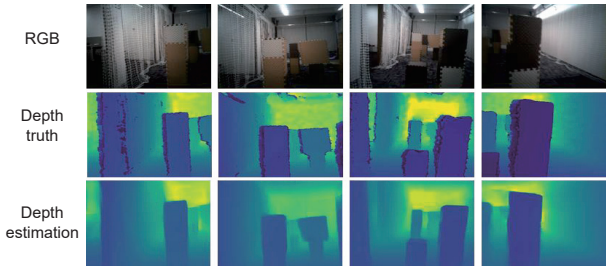

Fig. 2: Monocular depth estimation.

2) Monocular Depth Estimation: We developed a novel lightweight depth estimation framework that fuses sparse ToF depth data with RGB images (obtained from an OV5647 camera), leveraging geometrically precise ToF measurements simulated on BlendedMVS [17] and MegaDepth [18] datasets using 8×8 grid sampling, to resolve the inherent scale ambiguity of monocular depth estimation. An input-level fusion strategy is employed, where sparse 8×8 ToF depth is upsampled to RGB resolution and transformed into a 3D geometric feature map, fused with the normalized RGB image at the earliest processing stage to optimize multimodal alignment while reducing model complexity. This is followed by an efficient hierarchical encoder-decoder architecture employing a lightweight MobileNetV2 backbone (1.2 M parameters, ImageNet pre-trained) [19] for 32× downsampled feature extraction, with the decoder progressively upsampling features through bilinear interpolation and skip-connections to fuse multi-scale information up to 4x resolution. A convolutional residual refinement module subsequently enhances depth edges and outputs full-resolution predictions via final 4× upsampling, trained end-to-end with L1 loss. As shown in Fig. 2, the framework achieves distance measurement error of 0.47 m within a range of 6 m.

3) Monocular Direction Measurement: We developed a lightweight monocular vision system to achieve angle measurement, using an OmniVision OV5647 camera with a FOV of 72.9° and a resolution of 320×240 pixels. To identify the cooperative target in the image, a lightweight YOLOv7-tiny model [20] with 1.5 M parameters, 828 M FLOPs and 99.2% precision is trained and deployed on an onboard RISC-V AI processor, running at 15 Hz, as shown in Fig. 3. After recognizing the cooperative target, the coordinates of the target center in the two-dimensional image plane can be acquired. Then the relative direction between UAVs in the world coordinate system can be calculated according to the principle of monocular camera imaging. The developed relative direction finding technique yields an accuracy of 1.2° in verification experiments.

Fig. 3: Detecting UAVs using monocular camera.

4) Optoelectronic Detection: We designed a relative direction measurement system based on a 2D optoelectronic position-sensitive detector. This system detects optical beacons mounted on other UAVs within the swarm to measure relative bearings. The sensor integrates a micro-lens with a 72° field of view and employs a 660 nm filter for near-infrared cooperative target detection. As shown in Fig. 4, Gaussian centroid algorithm locates spot centroids on the image plane, while the pinhole camera model (similar triangles principle) calculates relative directions. Integrated with an ARM embedded computing unit, the system achieves a direction measurement accuracy of 0.9° at an update rate of 100 Hz.

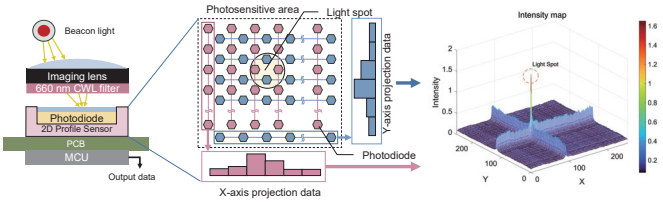

Fig. 4: The 2D optoelectronic sensor.

5) State Estimation: We developed a lightweight state estimation system consisting of a MEMS IMU, a ToF sensor to the ground and an optical-flow system. The Extended Kalman Filter (EKF) is utilized to fuse sensor outputs and generate smooth pose estimation results. Inertial pose measurement exhibits drift and error accumulation over time and flight distance. However, its estimations of velocity and acceleration remain highly accurate over brief periods, providing essential complementary references for relative positioning.

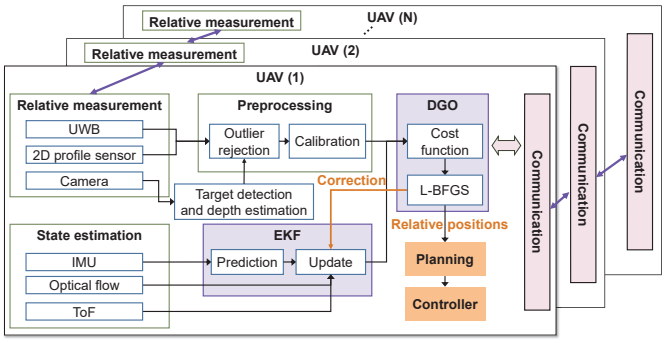

Fig. 5: Flowchart of the relative positioning framework.

B. Relative Positioning Based on DGO-EKF coupling

Based on the multimodal measurements, the relative positions are solved through DGO. The swarm's relative position topology map is decomposed into multiple subgraphs, enabling each UAV to solve its local topology on-board. Through a limited information sharing mechanism, the locally computed solutions iteratively converge to the globally optimal solution. Furthermore, the DGO results are fed as measurement inputs into the UAVs' pose estimation system using EKF. This corrects the accumulated errors in state estimation over time, as shown in Fig. 5.

The workflow of DGO is detailed as follows. Considering a swarm of N UAVs, we define the real position of UAV (i) at step k as $\mathbf{p}_{k,(i)} = \left[x_{k,(i)}, y_{k,(i)}, z_{k,(i)}\right]^T$. The real position set of the swarm can be written as $\mathbf{P}_k = \left\{\mathbf{p}_{k,(1)}, \mathbf{p}_{k,(2)}, \cdots, \mathbf{p}_{k,(N)}\right\}$. We give the distance measurement model as

$$\hat{d}_{k,(i,j)} = h^d\left(\mathbf{p}_{k,(i)}, \mathbf{p}_{k,(j)}\right) + \mathcal{N}\left(0, \Sigma_d^2\right) \tag{1}$$

where $\hat{d}_{k,(i,j)}$ represents the distance between UAV (i) and UAV (j) at step k, measured by the distance sensor mounted on UAV (i). h^d is the observation function of distance. $\mathcal{N}$ stands for the normal distribution function, and Σ_d is the variance of Gaussian noise matrix.

The measurement model of relative direction is given as

$$\hat{\theta}_{k,(i,j)} = h^\theta\left(\mathbf{p}_{k,(i)}, \mathbf{p}_{k,(j)}\right) + \mathcal{N}\left(0, \Sigma_\theta^2\right) \tag{2}$$

where $\hat{\theta}_{k,(i,j)}$ is the relative angle on the horizontal plane between UAV (i) and UAV (j) at step k, measured by the direction sensor mounted on UAV (i). h^θ is the observation function of direction, and Σ_θ is the variance of Gaussian noise matrix.

The UAVs solve the relative position graph in a distributed optimization scheme. At step k, UAV (i) regards its own position $\mathbf{p}_{k,(i)}$ as the variable to be optimized. Utilizing local relative measurements and the positions of other UAVs in the swarm ($\left\{\mathbf{p}_{k,(j)}, j \neq i\right\}$), a series of measurement residuals can be constructed as cost functions. Although the estimated $\mathbf{p}_{k,(j)}$ is unknown to UAV (i) at step k, UAV (i) can receive the positions of other UAVs estimated by themselves at step $k-1$ through inter-UAV communication with limited byte capacity, which can be expressed as $\left\{\hat{\mathbf{p}}_{k-1,(j)}, j \neq i\right\}$. To

compensate for the displacement between step $k-1$ and step k, we have

$$\mathbf{p}^{C}_{k,(j)} = \hat{\mathbf{p}}_{k-1,(j)} + \mathbf{v}^{pose}_{k-1,(j)}\Delta t, j \neq i \tag{3}$$

where $\mathbf{p}^{C}_{k,(j)}$ denotes the compensated position of UAV (j) at step k, and $\mathbf{v}^{pose}_{k-1,(j)}$ is the velocity of UAV (j) at step $k-1$ obtained by pose estimation system. The distance error function of UAV (i) can be expressed as

$$J^{d}_{k,(i)} = \sum_{j\in\mathbf{O}_d} \left\| \hat{d}_{k,(i,j)} - \left\| \mathbf{p}_{k,(i)} - \mathbf{p}^{C}_{k,(j)} \right\|_2 \right\|^2_{\Sigma_d} \tag{4}$$

where the set $\mathbf{O}_d$ stores number j, if UAV (j) is within the distance measurement range of UAV (i). $\|(*)\|^2_{\Sigma} = (*)^T\Sigma^{-1}(*)$ is the Mahalanobis norm of matrix (*). The angle error function can be also given as

$$J^{\theta}_{k,(i)} = \sum_{j\in\mathbf{O}_\theta} \left\| \hat{\theta}_{k,(i,j)} - \angle\left(\mathbf{p}_{k,(i)}, \mathbf{p}^{C}_{k,(j)}\right) \right\|^2_{\Sigma_\theta} \tag{5}$$

where $\mathbf{O}_\theta$ is the set of all UAVs in the view of the direction sensor mounted on UAV (i), $\angle(*,*)$ calculates the relative angle between two given coordinates. In real-world scenarios, owing to the limited FOV of the camera and optoelectronic sensor, $\mathbf{O}_\theta$ usually contains only a subset of the cooperative UAVs. However, since the UWB sensor is omnidirectional, the relative position graph of the swarm can be determined using sufficient distance information and a few pairs of relative angle measurements.

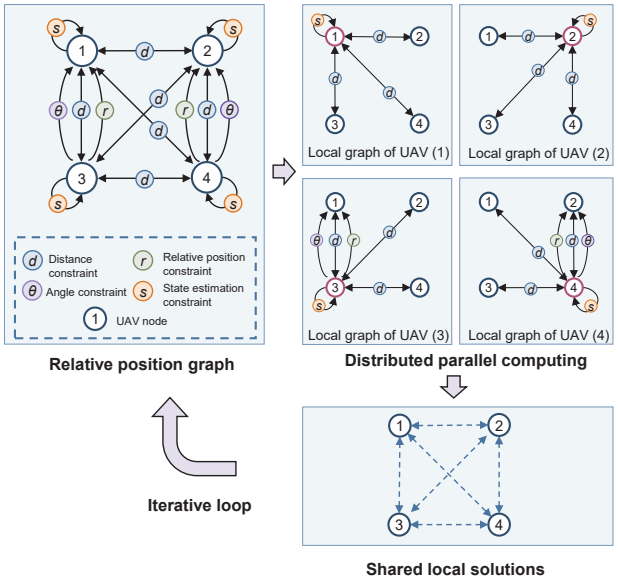

Fig. 6: Diagram of the DGO method.

In addition, the pose estimation system of UAV (i) provides high-frequency position measurements. Although the estimation error accumulates over time, the displacement measurement in a short period is relatively accurate. The error function of pose estimation is given as

$$J^{pose}_{k,(i)} = \left\| \Delta\hat{\mathbf{p}}^{pose}_{k-1\to k,(i)} - \left(\mathbf{p}_{k,(i)} - \hat{\mathbf{p}}_{k-1,(i)}\right) \right\|^2_{\Sigma_o} \tag{6}$$

where $\Delta\hat{\mathbf{p}}^{pose}_{k-1\to k,(i)}$ is the displacement of UAV (i) from step $k-1$ to step k, $\hat{\mathbf{p}}_{k-1,(i)}$ is the position of UAV (i) at step $k-1$, which is obtained through optimization at step k.

Utilizing the error functions given by (4)-(6), the graph optimization problem can be formulated as follows:

$$\hat{\mathbf{p}}_{k,(i)} = \underset{\mathbf{p}_{k,(i)}}{\arg\min} \left(J^{d}_{k,(i)} + J^{\theta}_{k,(i)} + J^{pose}_{k,(i)} \right) \tag{7}$$

Based on DGO, the relative positions of the swarm can be solved in a distributed scheme utilizing the L-BFGS method [21]. We further proposed a DGO-EKF coupling system that integrates low-frequency DGO estimates into the EKF as measurement updates. The UAVs uniformly designate one of the nodes in the swarm as the reference point. All UAVs in the swarm, excluding the reference point, utilize the results of DGO to update the Kalman gain, state variables, and covariance matrix. The proposed system constrains the dispersed UAVs near the reference node through relative measurements, preventing the accumulation errors of pose estimation over time.

III. Cooperative Target Localization

Leveraging the relative positioning scheme, the swarm unleashes collective intelligence capabilities, which is exemplified by cooperative target tracking missions.

We first developed a lightweight human pose recognition model based on the residual log-likelihood estimation (RLE) method [22]. MobileNetV2 with a width multiplier of 0.35 was utilized as the backbone [19]. The model was trained on the COCO dataset [23], which is augmented with Gaussian blur and contrast adjustments. The validation set comprised the COCO key point validation set and a COCO-formatted key point set consisting of 368 images captured by the onboard OV5647 camera of the UAV. After post-training quantization, the model was deployed on on the RISC-AI processor of the UAV, achieving a mean average precision (mAP) of 0.840.

For collaborative target localization, it is essential for UAVs to verify whether the detected targets are identical (the person ReID problem). In the swarm system, the key points of the targets extracted from the model on each UAV are shared among the UAVs. Then, each UAV performs local-to-received feature matching. After the ReID process, the relative pose between each UAV and the desired target is computed using the previously calibrated camera intrinsic parameters. The relative angles between the UAVs and the target in the horizontal plane are extracted as $\mathbf{A}^{\text{tgt}}_{k} = \left\{\alpha^{\text{tgt}}_{k,(1)}, \alpha^{\text{tgt}}_{k,(2)}, \cdots, \alpha^{\text{tgt}}_{k,(N)}\right\}$. Fusing the inter-UAV relative positions obtained from DGO calculations with the relative angles between the UAVs and the target, an overall pose graph can be generated, and the distributed optimization is performed. The cost function is defined as

$$J^{tgt}_{k} = \sum_{i\in\mathcal{O}_{tgt}} \left(\alpha^{tgt}_{k,(i)} - \angle\left(\hat{\mathbf{p}}_{k,(i)}, \mathbf{p}^{tgt}_{k}\right) \right)^2 \tag{8}$$

where $\mathbf{p}^{tgt}_{k}$ represents the position of the target. The set $\mathcal{O}_{tgt}$ stores number i, if the target can be detected by UAV (i).

$\mathbf{p}_k^{tgt}$ can be estimated by:

$$\hat{\mathbf{p}}_k^{tgt} = \arg\min_{\mathbf{p}_k^{tgt}} J_k^{tgt} \tag{9}$$

IV. Collaborative Target Tracking Based on Voronoi Diagram

Extending our prior framework for relative measurement and target localization, we implemented cooperative tracking and encirclement of dynamic targets through distributed Voronoi partitioning.

A. Problem Formulation

We consider a swarm consisting of N_p UAVs acting as pursuers and a target serving as the evader, constrained within a bounded dynamic convex environment $Q \subset \mathbb{R}^2$. The positions of pursuers are $\mathbf{p}_{i_p}^p = \left(x_{i_p}^p, y_{i_p}^p\right) \in Q, i_p = 1, 2, \cdots, N_p$, and the position of the evader is $\mathbf{p}^e = (x^e, y^e) \in Q$. Assume that the motion of both the pursuers and the evader can be described by the same dynamical equations:

$$\begin{aligned} &\dot{\mathbf{p}}_{i_p}^p(t) = \mathbf{v}_{i_p}^p(t), \mathbf{p}_{i_p}^p(0) = \mathbf{p}_{0,i_p}^p, i_p = 1, 2, \cdots, N_p \\ &\dot{\mathbf{p}}^e(t) = \mathbf{v}^e(t), \mathbf{p}^e(0) = \mathbf{p}_0^e \end{aligned} \tag{10}$$

where $\mathbf{v}_{i_p}^p(t)$ and $\mathbf{v}^e(t)$ represent the velocity inputs of the pursuers and the evader, respectively. $\mathbf{p}_{0,i_p}^p, \mathbf{p}_0^e \in Q$ are the initial positions of the pursuers and the evader. To ensure the successful encirclement of the target by the UAVs, it is assumed that their maximum speeds are the same:

$$\|\mathbf{v}_{i_p}^p(t)\|_2 \le v_{max}, \|\mathbf{v}^e(t)\|_2 \le v_{max} \tag{11}$$

In practical situations, the UAVs typically possess speeds that are greater or equal to those of the moving targets, that is, $\|\mathbf{v}_{i_p}^p(t)\|_2 \ge \|\mathbf{v}^e(t)\|_2$. To complete the encirclement, N_p hunting points ($\mathbf{p}_{i_h}^h = \left(x_{i_h}^h, y_{i_h}^h\right)$) are uniformly generated along a circumference with a radius of r_h, centered around the evader. The objective of target hunting is to enable the UAVs to reach the hunting points in the shortest time:

$$\|\mathbf{p}_{i_h}^p(t_c) - \mathbf{p}_{i_h}^h(t_c)\|_2 \le \hat{r}_c, i_h = 1, 2, \cdots, N_p \tag{12}$$

where t_c is the capture time, $\hat{r}_c$ is the manually defined capture radius.

B. Pursuit Using Voronoi Diagram

Consider a set of points $P = \left\{\mathbf{p}^e, \mathbf{p}_1^p, \mathbf{p}_2^p, \cdots, \mathbf{p}_{N_p}^p\right\}$ representing the positions of the pursuers and the evader. Define $V(Q) = \left\{V^e, V_1^p, V_2^p, \cdots, V_{N_p}^p\right\}$ as the standard Voronoi partition of the bounded dynamic convex environment Q, where V^e is the Voronoi Cell (VC) of the evader, and $V_{i_p}^p$ is the VC of pursuer i_p. $N_i = \{1, 2, \cdots, N_p\}$ is the set of UAV indices. The VCs are given by

$$\begin{aligned} V^e =& \left\{\mathbf{p} \in Q \middle| \|\mathbf{p} - \mathbf{p}^e\| \le \|\mathbf{p} - \mathbf{p}_{i_p}^p\|, \forall i_p \in N_i\right\} \\ V_{i_p}^p =& \left\{\mathbf{p} \in Q \middle| \|\mathbf{p} - \mathbf{p}_{i_p}^p\| \le \min\left(\|\mathbf{p} - \mathbf{p}^e\|, \|\mathbf{p} - \mathbf{p}_{j_p}^p\|\right), \right. \\ & \left. \forall i_p, j_p \in N_i, j_p \ne i_p\right\} \end{aligned} \tag{13}$$

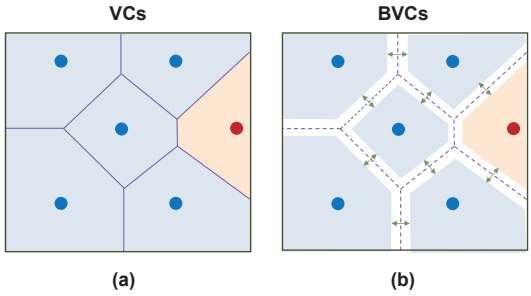

Fig. 7: The VCs and BVCs.

The mathematical interpretation of a Voronoi diagram involves partitioning the plane into multiple polygons, each containing exactly one generating point. The points within each polygon are closer to the corresponding generating point than to any other generating point. In the pursuit-hunting game, the target and UAVs are used as the generating points. V^e represents the safe-reachable area of the target, and $V_{i_p}^p$ represents the hunting area of pursuer i_p.

In a standard Voronoi diagram, the VCs share common edges. When the UAVs move within their respective VCs, there is a risk of inter-UAV collisions due to excessively small distances between UAVs. Based on the Voronoi diagram, the concept of buffered Voronoi cells (BVCs) is introduced. We define ${}^bV(Q) = \left\{{}^bV^e, {}^bV_1^p, {}^bV_2^p, \cdots, {}^bV_{N_p}^p\right\}$ as the buffered Voronoi partition of Q. The BVCs of P are given by

$$\begin{aligned} {}^bV^e =& \left\{\mathbf{p} \in Q \middle| \|\mathbf{p} - \mathbf{p}^e\| \le \|\mathbf{p} - \mathbf{p}_{i_p}^p\| - r_e, \forall i_p \in N_i\right\} \\ {}^bV_{i_p}^p =& \left\{\mathbf{p} \in Q \middle| \|\mathbf{p} - \mathbf{p}_{i_p}^p\| \le \min\left(\|\mathbf{p} - \mathbf{p}^e\|, \|\mathbf{p} - \mathbf{p}_{j_p}^p\|\right) \right. \\ & \left. - r_p, \forall i_p, j_p \in N_i, j_p \ne i_p\right\} \end{aligned} \tag{14}$$

where r_e and r_p represent the weights of the buffer distances between BVCs. Given a minimum safe distance d_s for inter-UAV collision avoidance, we have

$$\begin{aligned} r_e &= d_s\|\mathbf{p}^e - \mathbf{p}_{i_p}^p\|, i_p \in N_n \\ r_s &= \begin{cases} d_s\|\mathbf{p}_{i_p}^p - \mathbf{p}_{j_p}^p\|, j_p \in N_n \\ d_s\|\mathbf{p}_{i_p}^p - \mathbf{p}^e\|, e \in N_n \end{cases} \end{aligned} \tag{15}$$

where N_n is the set of UAV indices that are Voronoi neighbors of a UAV, and e is the index of the evader. The differences between the standard VCs and the BVCs are shown in Fig. 7.

C. Goal Point Generation

Based on the computed BVCs, we developed a distributed hunting strategy. The proposed strategy can be divided into two stages. Given a appropriate distance constant d_h, when $\|\mathbf{p}_{i_p}^p - \mathbf{p}^e\| \ge d_h$, we adopt a global area-minimization strategy to minimize the BVC area of the evader. Calculating the optimal goal points via the evader's BVC area differential is slow. A common approximation uses the midpoints of the shared edges between the evader's and pursuers' BVCs. If

there are no shared edges between a pursuer and the evader, the evader is set to be the goal points. When $\|\mathbf{p}_{i_p}^p - \mathbf{p}^e\| < d_h$, we consider that the pursuers have largely completed the hunting process. At this stage, the pursuers are set to fly directly towards the hunting points $\mathbf{p}_{i_h}^h$ to form an encirclement.

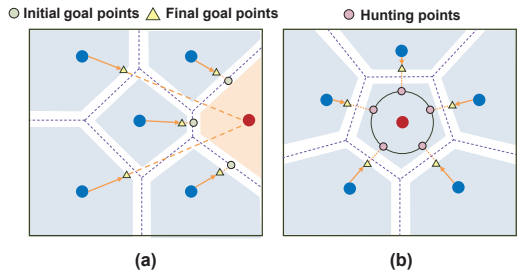

Fig. 8: Diagram of generating goal points.

We define $\mathbf{g}_{i_p}$ as the initial flight goal point of purser i_p. As illustrated in Fig. 8, the generated goal points may not be located within the BVCs. To avoid inter-aircraft collisions, we adjust the goal points by connecting the UAV and the goal point with a straight line, relocating the goal point to the intersection of this line and the BVC boundary, which is denoted as $\mathbf{g}_{i_p}^*$. Finally, each UAV inputs the locally computed $\mathbf{g}_{i_p}^*$ into the low-level control system to enable swarm control.

V. Experimental Results

A. Experimental Platform

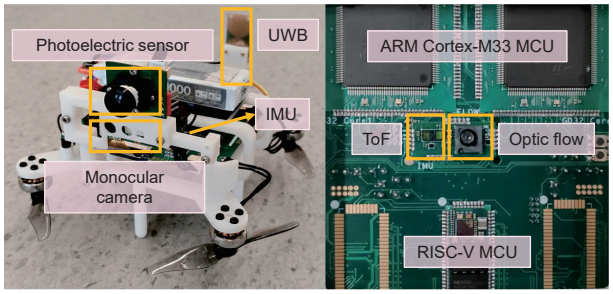

Fig. 9: The integrated micro-UAV platform.

To validate the proposed relative positioning and control framework, we integrated sensing, computing, and communication units into a microsystem, and formed a 150 g typical micro-UAV, as shown in Fig. 9. The sensors for relative measurement were introduced in Section II. For computation, we employed a RISC-V MCU to run lightweight neural networks on-device for processing visual information, and utilized two ARM Cortex-M33 MCUs to handle sensor data fusion, DGO computations, and control algorithms. Inter-UAV communication is achieved through UWB, transmitting data packets of less than 100 bytes at a rate of 10 Hz. This data includes local DGO solutions, control commands, among other information. The integrated core unit of the UAV operates at a power consumption of 2.3 W. Compared to existing high-computation vision-based UAV systems, the developed system effectively reduces weight and energy consumption.

B. Field Experiment of Autonomous Swarm Formation Flight

To validate the capability of the proposed framework to achieve autonomous flight without external infrastructure, we conducted an outdoor long-distance formation flight experiment using four micro-UAVs. The swarm flew 80 m at a speed of 1 m/s while maintaining a square formation. Relying solely on onboard multimodal sensors and the relative position-solving algorithm, the swarm achieved real-time measurement and calculation of their relative positions. We used GNSS-RTK modules to obtain the ground truth positions of the UAVs. It is important to note that the RTK modules did not participate in the solving process.

As shown in Fig. 10, we compared the proposed relative positioning framework with the method using only the UAVs' self-pose estimation. It is observed that the UAVs maintained stable square formation flight throughout the entire duration while operating the proposed relative positioning framework. In contrast, the UAVs exhibited significant positional drift when relying solely on their self-pose estimation, potentially leading to inter-UAV collisions.

We further conducted three sets of quantitative comparison experiments. The flight trajectories of the swarm and the relative positioning error curves are plotted in Fig. 11. Specifically, taking UAV1 as the reference, the relative positioning errors between UAV1 and UAV2, UAV3, and UAV4 were evaluated separately. The results indicate that in all three experimental sets, the relative positioning error of the self-pose estimation method shows a linear increase over time. This is attributed to system drift caused by the accumulation of measurement noise during the integration of inertial sensor data. When applying the proposed relative positioning framework, the estimated trajectory closely aligns with the ground truth, effectively mitigating dispersion drift within the cluster. This is because the proposed framework not only solves relative positions but also couples them with pose estimation feedback. This enables iterative correction of swarm dispersion drift, while featuring high precision and a lightweight design. As shown in TABLE I, across the three experiments, the average relative position errors calculated using the proposed method are 0.202 m, 0.338 m, and 0.247

TABLE I: Relative positioning error in formation flight

No.	Method	Relative positioning error (m)			
		UAV1-2	UAV1-3	UAV1-4	Average
1	Self-pose	3.065	3.355	4.264	3.561
	Proposed	0.302	0.158	0.147	**0.202**
2	Self-pose	1.910	4.524	6.073	4.169
	Proposed	0.278	0.301	0.435	**0.338**
3	Self-pose	1.30	1.359	3.150	1.938
	Proposed	0.184	0.209	0.346	**0.247**

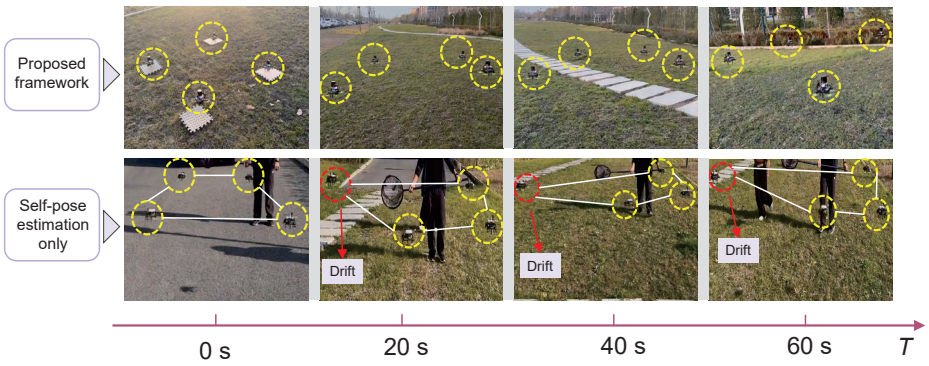

Fig. 10: Swarm formation flight photos, using the relative positioning framework and self-pose estimation only, respectively

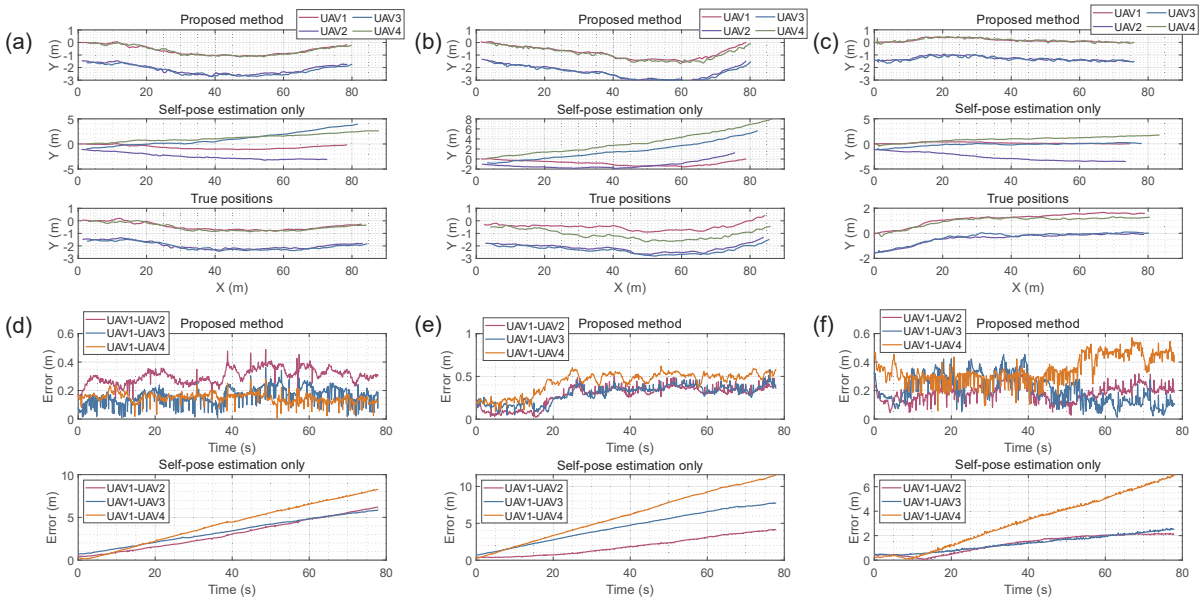

Fig. 11: Results of the formation flight experiments. (a)-(c) The flight trajectories in the three sets of experiments. (d)-(f) The relative positioning errors in the three sets of experiments.

m respectively, demonstrating stable error magnitudes. In contrast, the average relative position errors derived from the onboard pose measurement system are 3.561 m, 4.169 m, and 1.939 m. These field experiments validate the effectiveness of the relative position solving framework based on DGO.

C. Experiment of Swarm Autonomous Target Tracking

Combining the proposed relative positioning framework and the Voronoi-diagram-based distributed control scheme, we validated the swarm's control and decision-making capabilities through dynamic target tracking experiments. Six UAVs were utilized in the experiment. Three UAVs were responsible for locating the target, while the other three tracked it. The UAVs had a maximum speed set to 1.1 m/s and a maximum acceleration of 1.4 m/s^2. The human target object moved in an unstructured manner at a maximum speed of approximately 1.0 m/s.

The flight trajectories of the swarm during the cooperative tracking process are depicted in Fig. 12. During the initial phase (t=10 s to 12 s), the target maneuvered at its maximum speed of 1 m/s. At this point, the swarm rapidly formed a triangular interception formation using the Voronoi diagram planning method. Notably, at t=22 s, the target abruptly changed direction to execute a strategic evasion maneuver. The swarm re-planned its target points via the Voronoi diagram method, reconstructing the encirclement within 10 seconds, demonstrating excellent environmental adaptability. This entire process was repeated four times, successfully achieving stable tracking of the target.

The distance variation curves between the three UAVs and the target, shown in Fig. 12, reveal the core convergence characteristics of the swarm cooperative autonomous motion control method. The results indicate that during all four target captures, the distance curves between the UAVs and the target reached a convergent state, satisfying the predefined capture conditions. Operating within the dynamic system performance limits of a maximum speed of approximately 1.1 m/s and a maximum acceleration of approximately 1.4

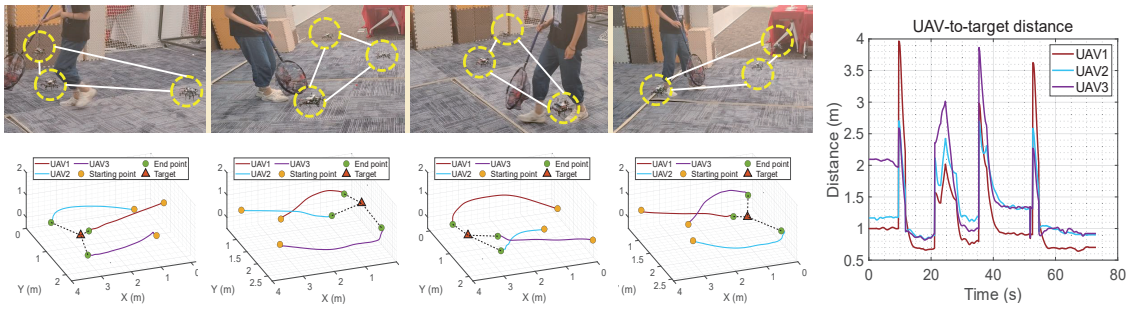

Fig. 12: Trajectories and distance curves in the cooperative target hunting process.

m/s², the swarm achieved stable tracking and capturing of the target moving at its maximum speed of 1 m/s.

VI. Conclusions

This paper presents an all-onboard multimodal relative positioning and control framework for hundred-gram-scale micro-UAV swarm. First, we propose a multimodal inter-UAV relative measurement system based on Vision-Optoelectronic-UWB fusion. Second, we develop a real-time and omnidirectional DGO framework for relative position computation. This solution is integrated with UAV pose estimation to enable iterative refinement of positioning accuracy. Finally, we integrate the determined relative positions into swarm cooperative target localization and implement swarm control using a distributed Voronoi diagram approach. Swarm flight experiments with 150 g micro-UAVs demonstrate that by utilizing the proposed framework, the swarm can autonomously complete long-distance formation flights without GNSS or external base stations, achieve decimeter-level relative positioning accuracy, and execute coordinated tracking and encirclement of dynamic targets. The proposed architecture is crucial for achieving autonomy in micro-UAVs and enabling the emergence of swarm intelligence.

References

[1] S. Javed, A. Hassan, R. Ahmad, W. Ahmed, R. Ahmed, A. Saadat, and M. Guizani, "State-of-the-art and future research challenges in uav swarms," *IEEE Internet of Things Journal*, vol. 11, no. 11, pp. 19 023–19 045, 2024.

[2] Y. Liu, Y. Wang, J. Wang, and Y. Shen, "Distributed 3d relative localization of uavs," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 11 756–11 770, 2020.

[3] C. Ren, Y. Jiao, Y. Liu, and H. Shang, "Optimal camera focal length detection method for gps-supported bundle adjustment in uav photogrammetry," *Measurement*, vol. 228, p. 114329, 2024.

[4] B. Yang, E. Yang, L. Yu, and A. Loeliger, "High-precision uwb-based localisation for uav in extremely confined environments," *IEEE Sensors Journal*, vol. 22, no. 1, pp. 1020–1029, 2021.

[5] P. Gu and Z. Meng, "S-vio: Exploiting structural constraints for rgb-d visual inertial odometry," *IEEE Robotics and Automation Letters*, vol. 8, no. 6, pp. 3542–3549, 2023.

[6] A. Luo, X. Li, F. Yang, J. Liu, H. Fan, and S. Liu, "Flowdiffuser: Advancing optical flow estimation with diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2024, pp. 19 167–19 176.

[7] A. Rajvanshi, H.-P. Chiu, A. Krasner, M. Sizintsev, G. Murray, and S. Samarasekera, "Ranging-aided ground robot navigation using uwb nodes at unknown locations," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 786–793.

[8] Y. Tang, Y. Hu, J. Cui, F. Liao, M. Lao, F. Lin, and R. S. Teo, "Vision-aided multi-uav autonomous flocking in gps-denied environment," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 1, pp. 616–626, 2018.

[9] X. Oh, R. Lim, L. Loh, C. H. Tan, S. Foong, and U.-X. Tan, "Monocular uav localisation with deep learning and uncertainty propagation," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7998–8005, 2022.

[10] G. Zhang and L.-T. Hsu, "Intelligent gnss/ins integrated navigation system for a commercial uav flight control system," *Aerospace science and technology*, vol. 80, pp. 368–380, 2018.

[11] M. Basiri, F. Schill, P. Lima, and D. Floreano, "On-board relative bearing estimation for teams of drones using sound," *IEEE Robotics and Automation letters*, vol. 1, no. 2, pp. 820–827, 2016.

[12] B. Yang, E. Yang, L. Yu, and C. Niu, "Adaptive extended kalman filter-based fusion approach for high-precision uav positioning in extremely confined environments," *IEEE/ASME Transactions on Mechatronics*, vol. 28, no. 1, pp. 543–554, 2022.

[13] H. Xu, Y. Zhang, B. Zhou, L. Wang, X. Yao, G. Meng, and S. Shen, "Omni-swarm: A decentralized omnidirectional visual–inertial–uwb state estimation system for aerial swarms," *IEEE Transactions on Robotics*, vol. 38, no. 6, pp. 3374–3394, 2022.

[14] J. Jia, X. Chen, W. Wang, and M. Zhang, "Distributed control of target cooperative encirclement and tracking using range-based measurements," *Asian Journal of Control*, vol. 25, no. 6, pp. 4595–4608, 2023.

[15] H. Jafarzadeh and C. Fleming, "Dmpc: A data-and model-driven approach to predictive control," *Automatica*, vol. 131, p. 109729, 2021.

[16] M. Zhou, Z. Wang, J. Wang, and Z. Cao, "Multi-robot collaborative hunting in cluttered environments with obstacle-avoiding voronoi cells," *IEEE/CAA Journal of Automatica Sinica*, 2024.

[17] Y. Yao, Z. Luo, S. Li, J. Zhang, Y. Ren, L. Zhou, T. Fang, and L. Quan, "Blendedmvs: A large-scale dataset for generalized multi-view stereo networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[18] Z. Li and N. Snavely, "Megadepth: Learning single-view depth prediction from internet photos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[19] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[20] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 7464–7475.

[21] M. M. Najafabadi, T. M. Khoshgoftaar, F. Villanustre, and J. Holt, "Large-scale distributed l-bfgs," *Journal of Big Data*, vol. 4, no. 1, p. 22, 2017.

[22] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, and M. Shah, "Deep learning-based human pose estimation: A survey," *ACM computing surveys*, vol. 56, no. 1, pp. 1–37, 2023.

[23] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.