

TransformLoc: Transforming MAVs into Mobile Localization Infrastructures in Heterogeneous Swarms

Haoyang Wang¹, Jingao Xu², Chenyu Zhao¹, Zihong Lu³, Yuhang Cheng¹, Xuecheng Chen¹,
Xiao-Ping Zhang¹, Yunhao Liu², Xinlei Chen^{1,4,5†}

[†] Corresponding author

¹ Shenzhen International Graduate School, Tsinghua University, China;

² School of Software, Tsinghua University, China; ³ Harbin Institute of Technology, China;

⁴ Pengcheng Laboratory, Shenzhen, China; ⁵ RISC-V International Open Source Laboratory, Shenzhen, China

Email: {wanghaoyang0428, xujingao13, zhaocyhi, luzong2001, yhncc, 1243566go}@gmail.com,

xiaoping.zhang@sz.tsinghua.edu.cn, yunhao@tsinghua.edu.cn, chen.xinlei@sz.tsinghua.edu.cn

Abstract—A heterogeneous micro aerial vehicles (MAV) swarm consists of resource-intensive but expensive advanced MAVs (AMAVs) and resource-limited but cost-effective basic MAVs (BMAVs), offering opportunities in diverse fields. Accurate and real-time localization is crucial for MAV swarms, but current practices lack a low-cost, high-precision, and real-time solution, especially for lightweight BMAVs. We find an opportunity to accomplish the task by transforming AMAVs into mobile localization infrastructures for BMAVs. However, turning this insight into a practical system is non-trivial due to challenges in location estimation with BMAVs’ unknown and diverse localization errors and resource allocation of AMAVs given coupled influential factors. This study proposes TransformLoc, a new framework that transforms AMAVs into mobile localization infrastructures, specifically designed for low-cost and resource-constrained BMAVs. We first design an error-aware joint location estimation model to perform intermittent joint location estimation for BMAVs and then design a proximity-driven adaptive grouping-scheduling strategy to allocate resources of AMAVs dynamically. TransformLoc achieves a collaborative, adaptive, and cost-effective localization system suitable for large-scale heterogeneous MAV swarms. We implement TransformLoc on industrial drones and validate its performance. Results show that TransformLoc outperforms baselines including SOTA up to 68% in localization performance, motivating up to 60% navigation success rate improvement.

I. INTRODUCTION

Heterogeneous micro aerial vehicles (MAV) swarms offer transformative potential in conducting 4D (deep, dull, dangerous, dirty) tasks, especially quick response scenarios, such as search and rescue [1], gas leak source detection [2], wildfire suppression [3], leveraging their inherent advantages of working scalability [4], flexibility [5], adaptability [6], etc. There are forecasts that the market size for MAV swarm-supported applications will reach \$ 279 billion by 2032 [7].

A heterogeneous MAV swarm typically consists of: (i) a handful of advanced MAVs (AMAVs) equipped with intensive capabilities (eg. sensing, computing, etc.) yet expensive [8]; and (ii) a larger group of resource-limited and cost-effective basic MAVs (BMAVs) [9], as shown in Fig. 1¹. The

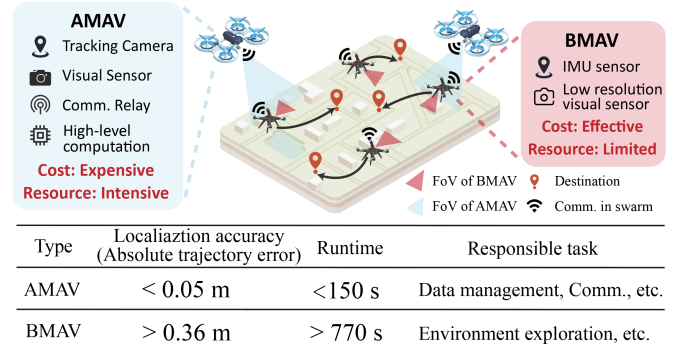


Fig. 1. Introduction of heterogeneous MAV swarm. The AMAVs are resource-intensive and have a lower localization error and latency, while BMAVs are resource-limited, resulting in a high error and latency.

former primarily handles data management, communication, and complicated computation tasks, while the latter is dispersed into (hazard) target areas for data collection and environmental exploration [12]. The AMAV-BMAV collaboration paradigm achieves an optimal balance between overall capabilities and cost, facilitating its widespread adoption [13].

However, the diverse onboard resources also lead to unbalanced localization performance between AMAVs and BMAVs, which is the fundamental capability for flight control [14], obstacle avoidance [15], etc. Precisely, localization errors of the BMAV accumulate fast due to its noisy sensing data and low computational capability, preventing it from achieving accurate and real-time localization as the AMAV.

Unfortunately, current methods are not able to offer feasible solutions for BMAV localization in quick response scenarios, which can be divided into 2 categories:

- **Extra infrastructure based solutions.** These solutions utilize external devices to provide reference signals (e.g., GPS [16], [17], RTK, radio [14], [18], or acoustic [19]) for localization. They require 1) pre-deploying (densely) expensive localization infrastructure in the operation site and 2) line-of-sight connection between localization infrastructure and MAVs. However, in quick response scenarios like urban disaster relief, either requirement is hard to satisfy, which leads to localization failures [20], [21].

¹Results in Fig.1 and Fig.2a are measured with MH_02 of EuRoC dataset [10] and CCM-SLAM [11]. AMAV is equipped with Intel(R) i7-8750H and 32G RAM, and BMAV is equipped with Cortex-A53 and 1G RAM.

• **Intra on-board sensor based solutions.** These solutions leverage onboard sensors like cameras, IMU, LiDAR, and Radar, along with simultaneous localization and mapping (SLAM) techniques, to achieve high-precision location estimation [22], [23]. These methods require intensive on-board sensing and computing capabilities, which only work for AMAVs. In contrast, BMAVs’ limited on-board resources result in significant accumulated errors or computation delays as illustrated in Fig.2a [15].

Therefore, this paper aims to improve BMAVs’ localization accuracy and efficiency given limited onboard sensing and computing capabilities without relying on any pre-deployed localization infrastructures. Our **key insight** is to transform a handful of AMAVs as *mobile localization infrastructures* and offload their sensing and computing capabilities to support location estimation improvement of a larger group of BMAVs. Specifically, AMAVs provide external observations (with their visual sensors) for BMAVs to correct cumulative location estimation errors [24], [25].

However, turning this insight into a practical system is non-trivial, since two technical challenges have to be solved:

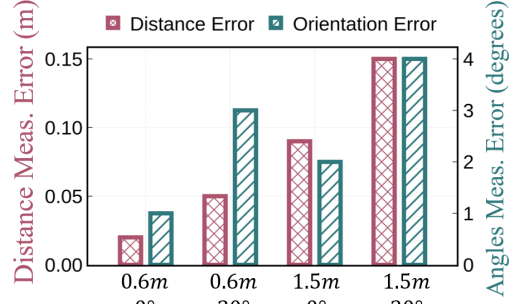
• **Unknown & Diverse localization errors of BMAVs (C1).** On the one hand, accumulated location estimation errors may be diverse among BMAVs due to various reasons, such as various sensing noises, different operational conditions, etc. On the other hand, due to its limited field of view (FoV), each AMAV can only serve several BMAVs as localization infrastructure. Therefore, AMAVs should provide external observations for BMAVs with significant errors for estimation correction. However, online deriving BMAVs’ localization errors is challenging due to the lack of static location references. As a result, localization errors of BMAVs may keep accumulating without timely external observations from AMAVs. Meanwhile, we need to consider how to correct BMAVs’ location with AMAV’s observations.

• **Resource allocation given coupled influential factors (C2).** Assigning sensing and computing resources of a handful of AMAVs to a larger group of BMAVs as localization infrastructure involves optimization in high dimensional decision space. First, even if C1 is solved, the presence of multiple AMAVs results in an exponential search space growth. Second, since both AMAVs and BMAVs are in constant motion, the observed BMAVs by each AMAV keep changing, necessitating dynamic adjustment of resource allocation strategies. Third, the localization effectiveness is affected by the accuracy of AMAVs’ observations given various distances and bearing angles between AMAVs and BMAVs (as shown in Fig. 2b), further complicating the resource allocation search space. These factors influence resource allocation decisions in a coupled way, making it too complicated to run on MAVs.

To conquer these challenges, this paper presents TransformLoc, a novel collaborative and adaptive location estimation framework for heterogeneous MAV swarms. TransformLoc dynamically transforms AMAVs with intensive onboard capabilities into mobile localization infrastructures to support

Algorithm	Drift	Latency
CCM-SLAM	< 0.3 cm	> 5 s
Dead-Reckoning	> 9 cm	< 0.05 s

(a) Accuracy & latency on BMAV



(b) Accuracy of AMAV observation

Fig. 2. Motivating example. (a) BMAVs have limited on-board resources, resulting in a trade-off between accuracy and latency; (b) The accuracy of observation is affected by distances and angles between MAVs.

location estimation improvement for resource-constrained BMAVs. To achieve accurate and real-time localization for the entire swarm, TransformLoc only requires deploying intensive and expensive sensing and computing capabilities on a small number of AMAVs, without relying on external localization infrastructures. Consequently, this framework facilitates the widespread deployment of swarm technology.

To address C1, we design an *error-aware joint location estimation model*. This model designs an uncertainty-aided inference method to enable AMAVs to provide observations for BMAVs with larger errors at first. Subsequently, it integrates inaccurate estimation from BMAVs with discontinuous external observations from AMAVs to perform intermittent joint location estimation for BMAVs.

To address C2, a *proximity-driven adaptive grouping-scheduling strategy* is proposed. Initially, MAVs are dynamically grouped according to the principle of proximity, transforming the many-to-many resource allocation problem into multiple one-to-many resource allocation problems. Then, several steps lookahead about BMAVs are conducted to schedule each AMAV in a non-myopic manner, which also finds optimal distance and angle.

We evaluate the performance of TransformLoc and compare it with baselines including the state-of-the-art (SOTA) method through extensive experiments with both real-time testbed and physical feature-based simulations. Results show that TransformLoc is able to maintain BMAVs’ average localization error under 1m with limited on-board resources in real-time, outperforming baselines up to 68%, which motivates up to 60% navigation success rate improvement.

In summary, the main contributions are as follows:

- We propose TransformLoc, a new framework that dynamically transforms AMAVs into mobile localization infrastructures, enhancing localization accuracy and real-time performance for lightweight BMAVs.

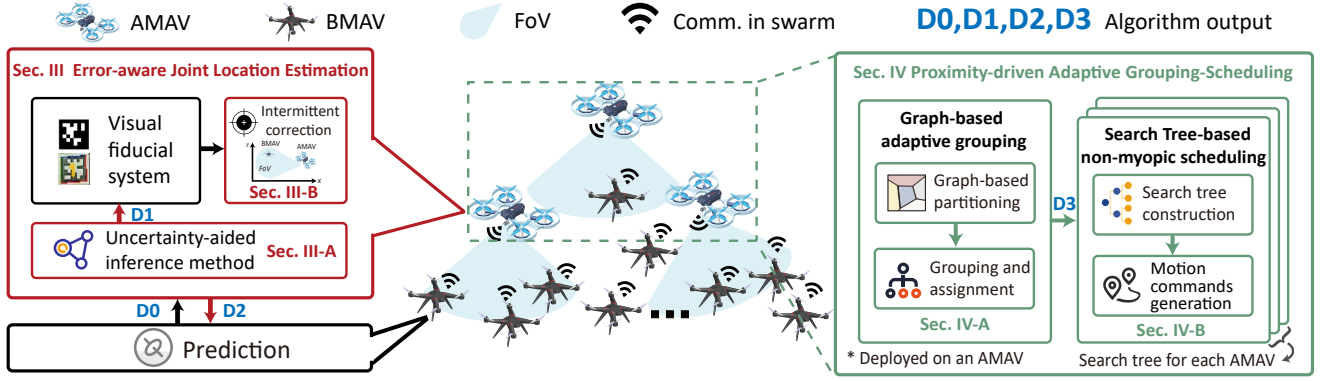


Fig. 3. Illustration of TransformLoc framework. BMAVs estimate locations with noisy measurements. With the assistance of the uncertainty-aided inference method, AMAV generates discontinuous observations for BMAVs to perform intermittent joint location estimation. Subsequently, TransformLoc allocates resources of AMAVs by adaptive grouping and scheduling, which adaptively groups MAVs at first, and then schedules AMAVs in a non-myopic manner.

- We design an *error-aware joint location estimation model* to boost the location estimation accuracy of BMAVs with discontinuous observation from AMAVs.
- We design a *proximity-driven adaptive grouping-scheduling strategy* to decouple the resource allocation issue given coupled influential factors.
- We validate our solution through in-field experiments on a real heterogeneous MAV swarm and large-scale physical feature-based simulations.

The remainder of the paper is organized as follows: Section II provides an overview of TransformLoc, with detailed descriptions of the error-aware joint location estimation model in Section III and the proximity-driven adaptive grouping-scheduling strategy in Section IV. Section V showcases the implementation and evaluation. Sections VI and VII discuss the related work and influencing factors of the framework, respectively. Section VIII concludes TransformLoc. Appendix sections include detailed formula derivations of essential variables.

II. OVERVIEW

A. TransformLoc: Framework goal

From the top perspective, we design and implement TransformLoc to transform AMAVs into mobile localization infrastructures for BMAVs. This adaptation is specifically crafted to cater to the economic and resource constraints of BMAVs, effectively addressing the inherent challenges presented by these limitations. The goal of framework design is to answer the questions:

- How to optimize the location estimation of BMAVs with unknown and diverse localization errors? TransformLoc should infer the localization error of BMAV and utilize observations generated by AMAV to assist BMAVs, effectively reducing the localization error of BMAV.
- How to allocate sensing resources of AMAVs given coupled influential factors to assist BMAVs? TransformLoc should decouple the resource allocation problem and navigate the AMAVs in a non-myopic way, ensuring that the overall localization error of BMAVs remains low.

B. Framework overview

The architecture of TransformLoc is illustrated in Fig. 3. As seen, TransformLoc consists of two main components:

- *Error-aware joint location estimation model*. Firstly, the BMAVs utilize the state model and motion commands to conduct the prediction of location at first, then transmit the yielded prior distribution of location to AMAVs ($D0$ in Fig.3). Secondly, the uncertainty-aided inference method (Sec. III-A) identifies when assistance is required for BMAVs ($D1$ in Fig.3). Finally, utilizing observations generated by visual fiducial system, AMAV intermittently corrects the location estimation of BMAVs (Sec. III-B), then transmits the result posterior distribution of BMAVs' location and motion commands to BMAVs for following motions ($D2$ in Fig.3).
- *Proximity-driven adaptive grouping-scheduling strategy*. The number of AMAVs is limited, and the proximity-driven adaptive grouping-scheduling strategy is responsible to allocate resources of AMAVs to BMAVs. Firstly, this strategy groups MAVs adaptively utilizing a graph-based adaptive grouping method according to the relationship of AMAVs and BMAVs in the spatial dimension (Sec. IV-A). Secondly, utilizing grouping result ($D3$ in Fig.3), TransformLoc schedules each AMAV to allocate the sensing resource in a non-myopic manner by constructing a search tree for each AMAV. Finally, TransformLoc generates motion commands for AMAVs to allocate sensing resources for BMAVs.

III. ERROR-AWARE JOINT LOCATION ESTIMATION MODEL

The BMAVs have unknown and diverse localization errors, complicating their location estimation with AMAV's assistance. In this part, we design an error-aware joint location estimation model based on Kalman filter and focus on localization error inference and joint location estimation of BMAVs. The main process is as follows:

- In order to enable AMAVs to provide observations for BMAVs with higher errors, we incorporate a measure of uncertainty to reflect the quality of BMAVs' location estimation in Sec III-A.
- Meanwhile, we estimate the location of BMAVs with two coupled operations in Sec III-B: *Prediction from state model* utilizes noisy motion measurements and the state model of

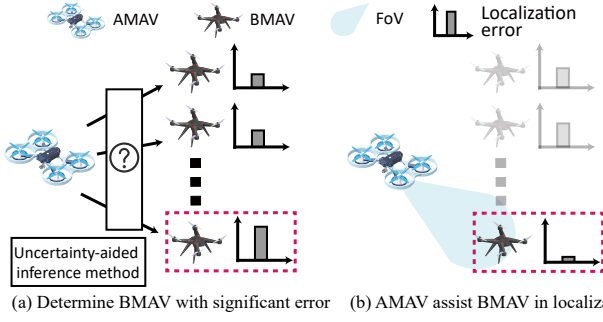


Fig. 4. The error-aware joint location estimation model. (a) AMAV determines the BMAV with a higher error under the assistance of the uncertainty-aided inference method, and (b) generates observations for correction.

BMAVs; *Correction from observation* incorporates linearized observations generated by AMAVs.

The state model of BMAV and AMAV, observation model of AMAV are described in Appendix A.

A. Uncertainty-aided inference method

The goal of TransformLoc is to efficiently allocate the sensing resources of AMAVs, thereby generating observations for BMAVs to mitigate localization errors. The localization error of BMAV B_i at time t can be mathematically expressed as

$$\psi(i, t) := E \left[\|y_{i,t} - \hat{y}_{i,t}\|^2 \right], \quad (1)$$

where $y_{i,t}$ is actual location of B_i and $\hat{y}_{i,t}$ is location estimation of B_i . By integrating the Kalman filter-based model for BMAVs' location estimation, we employ a metric of estimation uncertainty to gauge the accuracy of BMAVs' location estimations. This approach eliminates the necessity of having the actual locations of BMAVs to calculate the localization error. In order to measure the BMAV's estimation uncertainty, we choose the trace of the covariance matrix of BMAV's location estimation [26], which can be mathematically expressed as

$$\psi(i, t) = \text{tr}(\Sigma_{i,t}), \quad (2)$$

where $\Sigma_{i,t}$ is trace of the covariance matrix of B_i 's location estimation at time t . This indicator measures the uncertainty of estimations, a lower value indicates greater certainty. For BMAVs exhibiting varying degrees of localization errors (Fig. 4a), the AMAV generates observations tailored to assist BMAVs with more substantial errors (Fig. 4b).

B. Joint estimation of BMAVs' location

The joint location estimation framework of BMAVs includes two operations: ① prediction from the state model and noisy velocity measurement of BMAVs, which calculates the prior distribution of BMAVs' location; ② correction from the observation of AMAVs, which calculates the posterior distribution of BMAVs' location. The detail of joint estimation of BMAV's location is illustrated in Algorithm 1.

Prediction of BMAVs from state model. This process is outlined in lines 1-3 of Algorithm 1. It leverages the motion model of BMAVs, incorporating the noisy motion

Algorithm 1 A_j assists B_i for localization using the noisy motion measurement and the observation.

Input: Location estimation of B_i at time $t-1$, $\hat{y}_{i,t-1}$; covariance matrix of estimation at time $t-1$, $\Sigma_{i,t-1}$; noisy motion measurement, $\dot{v}_{i,t-1}$; location of A_j , $x_{j,t-1}$; motion command of A_j , $u_{j,t-1}$.

Output: Location estimation of B_i at time t , $\hat{y}_{i,t}$; covariance matrix of estimation at time t , $\Sigma_{i,t}$.

% Prediction of BMAV from state model

- 1: Update prior distribution of B_i 's location at time t ,
 $y_{i,t}^- = \int p(y_{i,t} | y_{i,t-1}, \dot{v}_{i,t-1}) \hat{y}_{i,t-1} dy_{i,t-1}$;
- 2: Update covariance of estimation $\Sigma_{i,t-1}^-$ from $y_{i,t}^-$;
- 3: Update $\hat{y}_{i,t}$ and $\Sigma_{i,t}$, from $y_{i,t}^-$ and $\Sigma_{i,t}^-$;
- % Correction of BMAV from observations*
- 4: Update the FoV of A_j , $F_{j,t}$ according to Eq.(5);
- 5: **if** B_i in the FoV of A_j **then**
- 6: Update observation $z_{i,j,t}$ according to Eq.(6);
- 7: Linearize the observation according to Eq.(7);
- 8: Update posterior distribution of B_i ' location at t ,
 $y_{i,t}^+ = \eta p(z_{i,j,t} | y_{i,t}^- y_{i,t}^-)$;
- 9: Update covariance of estimation $\Sigma_{i,t}^+$ from $y_{i,t}^+$;
- 10: Update $\hat{y}_{i,t}$ and $\Sigma_{i,t}$ from $y_{i,t}^+$ and $\Sigma_{i,t}^+$;
- 11: **end if**

measurement $\dot{v}_{i,t-1}$ and the location estimation $\hat{y}_{i,t-1}$ of BMAV B_i at time $t-1$ to compute the prior distribution and covariance matrix of B_i 's location estimation, denoted as $y_{i,t}^-$ and $\Sigma_{i,t}^-$.

Correction of BMAVs from observations. This process is delineated in lines 4-11 of Algorithm 1. When BMAV B_i is within the Field of View (FoV) of AMAV A_j at time t , this process employs the observation $z_{i,j,t}$ and the prior distribution of B_i 's location estimation to compute the posterior distribution and covariance matrix at time t , denoted as $y_{i,t}^+$ and $\Sigma_{i,t}^+$. The variable η in line 8 represents a normalization constant. If A_j generates an observation for B_i , the location estimation $\hat{y}_{i,t}$ and covariance matrix of estimation $\Sigma_{i,t}$ for B_i at time t are drawn from the posterior distribution $y_{i,t}^+$ and $\Sigma_{i,t}^+$; otherwise, they are drawn from the prior distribution $y_{i,t}^-$ and $\Sigma_{i,t}^-$.

IV. PROXIMITY-DRIVEN ADAPTIVE GROUPING-SCHEDULING STRATEGY

The resource allocation of AMAVs is influenced by coupled factors, involving optimization in a high-dimensional decision space. In this part, we design a *proximity-driven adaptive grouping-scheduling* strategy to allocate resources of AMAVs to assist BMAVs in localization. The main process is as follows:

- This strategy first dynamically groups the AMAVs and BMAVs according to the proximity in spatial domain based on the Voronoi diagram. This step transforms the many-to-many resource allocation problem into multiple one-to-many resource allocation problems.

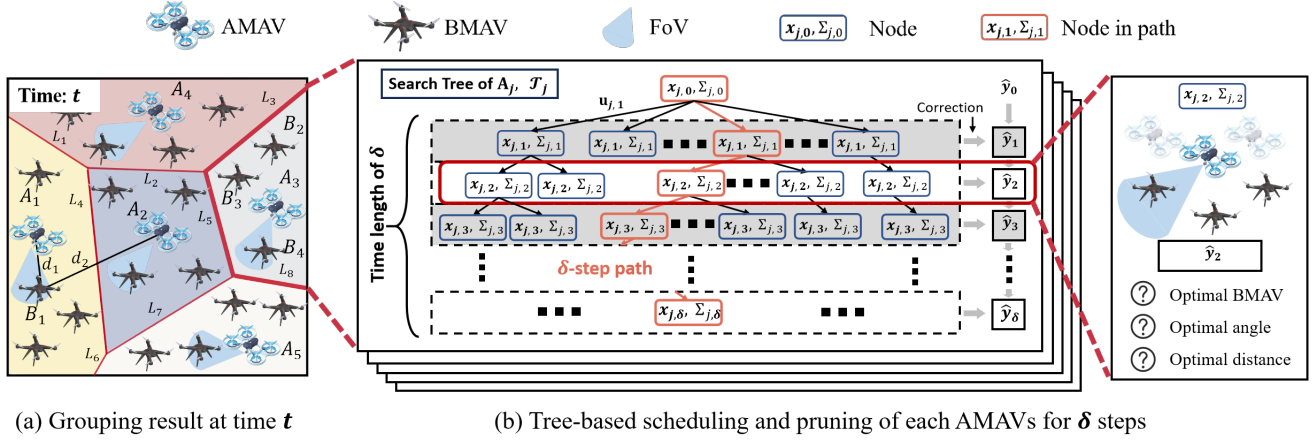


Fig. 5. The proximity-driven adaptive grouping-scheduling strategy. This strategy groups MAVs based on the Voronoi diagram to decouple the resource allocation problem. Then, it constructs search trees for each AMAV by involving several steps lookahead about BMAVs, resulting in δ -step paths for AMAVs. AMAV is scheduled in a non-myopic way to assist BMAVs with significant errors in an optimal distance and angle. $x_{j,0}$ is state of A_j , $\Sigma_{j,0}$ is covariance of assigned BMAVs' estimation maintained by A_j , and $\hat{y}_0$ is BMAVs' estimation location.

- Following that, to strategically plan each AMAV in a non-myopic manner and determine the optimal observational distance and angle, this strategy constructs a search tree for each AMAV, incorporating several steps of lookahead regarding BMAVs.

The mathematical formulation of the resource allocation problem is described in Appendix B.

A. Graph-based adaptive grouping

The BMAVs are located in various locations with varying localization errors. When generating observations for BMAVs, a single AMAV is limited by its location and waste sensing resources by moving between different BMAVs. In this section, we dynamically group BMAVs so that each AMAV can focus its sensing resources on one group. Grouping BMAVs and assigning them to different AMAVs pose a combinatorial optimization challenge that becomes inherently difficult to solve with a substantial number of BMAVs and AMAVs, owing to its NP-hard nature. To tackle this challenge, TransformLoc employs two key operations:

- TransformLoc initially divides the entire area into non-overlapping regions according to the locations of AMAVs.
- Subsequently, each AMAV allocates sensing resources for BMAVs within the nearest region for a duration of δ , which is the control command interval for BMAVs.

As a result, all of BMAVs are categorized into non-overlapping groups, with distinct groups being assigned to different AMAVs. The details are provided below.

Graph-based region partitioning. The partitioning approach is based on the Voronoi diagram [27]. Under this scheme, each AMAV is associated with a region encompassing points whose distance to the given AMAV is less than or equal to their distance to any other AMAV. Fig.5a presents a depiction of this partitioning method with five AMAVs and several BMAVs. The result is represented by region boundaries denoted as $L1 - L8$, which are separated by perpendicular bisectors of neighboring AMAVs. Any BMAV

located within the region of A_1 is closer to it than to any other AMAV (i.e., $d1 < d2$ in Fig.5a).

Grouping and assignment of MAVs. BMAVs within the designated region of an AMAV are organized into groups, and the AMAV allocates its sensing resources exclusively to assist in their localization for a duration of δ . The group renews after an interval of δ . Importantly, the boundaries of an AMAV's region are solely determined by the locations of its neighboring AMAVs and can be computed by identifying the perpendicular bisectors between adjacent AMAVs. If no BMAVs are present within an AMAV's region, it allocates sensing resources to all BMAVs over the duration of δ .

B. Search tree-based non-myopic scheduling

In this section, we present a methodology for integrating a search tree-based scheduling strategy into the non-myopic resource allocation of AMAVs. The primary procedure unfolds as follows:

- BMAVs receive commands at discrete time intervals of δ , enabling the acquisition of BMAVs' motion commands within each δ interval for AMAVs.
- These motion commands are subsequently employed to plan trajectories for AMAVs, incorporating a δ -step lookahead in coordination with the movements of BMAVs.

Search tree construction. We construct search trees for each AMAV. As Fig. 5b shows, we construct a search tree $\mathcal{T}_j$ for A_j as an example. $\mathcal{T}_j$ includes a set of candidate trajectories A_j can take, starting from an initial location and covariance pair $(x_{j,0}, \Sigma_{j,0})$, where $x_{j,0}$ is the starting location of A_j , and $\Sigma_{j,0}$ is the initial covariance of the location estimations of the BMAVs assigned to A_j . The nodes of the search tree at level $t \leq \delta$ correspond to reachable locations for A_j and are denoted as $(x_{j,t}, \Sigma_{j,t})$. The AMAV measures the distance and angle for observable BMAVs at each location to determine an optimal distance and angle to generate observations for BMAVs with significant error. We discretized the control space of the AMAV, A_j has a finite set of control options $\mathcal{U}$, with an edge for

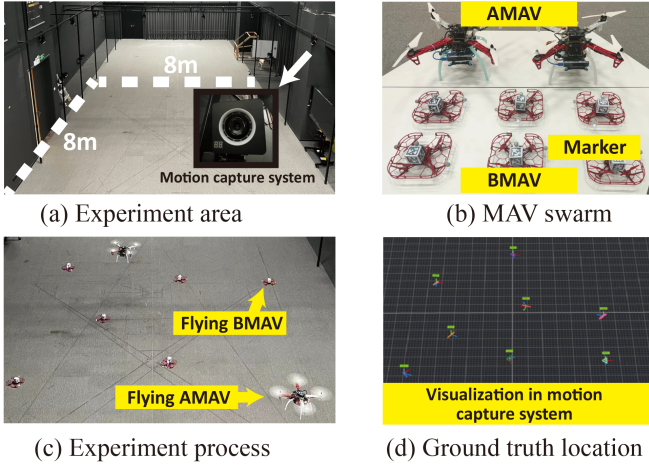


Fig. 6. Experiment area and heterogeneous MAV swarm implementation. The AMAVs utilize a tracking camera to localize themselves and visual sensors to generate observation. BMAVs mount AprilTag markers for recognition and localization. The ground truth is obtained from the motion capture system. The experiment area has a size of $8m \times 8m$.

each option $u_{j,t}$ starting at node $(x_{j,t}, \Sigma_{j,t})$ and leading to node $(x_{j,t+1}, \Sigma_{j,t+1})$ by evaluating state model of AMAV and Algorithm 1. Then estimation location of corresponding BMAVs $\hat{y}_t$ is computed by evaluating motion model of BMAV, observation model of AMAV, and Algorithm 1.

Motion commands for AMAV. Upon finishing the construction of the search tree for an AMAV A_j , we choose the node at level δ with the minimum value of $tr(\Sigma_{j,\delta})$. Subsequently, through backtracking on this node, we plan the trajectory of A_j navigate it for a duration of δ .

C. Scheduling of BMAV

TransformLoc utilizes location estimation to navigate BMAVs to their destinations. To achieve this, we design a lightweight planning algorithm based on an artificial potential field, ensuring BMAVs can avoid collision. The key points are as follows:

- TransformLoc maintains BMAVs' location estimation and generates motion commands with an interval of δ based on the distance to their destination.
- BMAVs move within a field of forces, the destination attracting them through a force proportional to the distance. The Wall and other BMAVs generate repulsive forces which repel the BMAV. This approach enables BMAVs to dynamically adjust motion when nearing the destination, decreasing velocity and increasing navigation success.

V. EVALUATION

A. Implementation and Methodology

Testbed Implementation. As Fig. 6 illustrated, we implemented the TransformLoc based on DJI Robomaster TTs (BMAVs) and industry drones (AMAVs) built on Pixhawk which is one of the most widely used autopilot systems, to validate it in the real world. The AMAV is equipped with an Intel(R) T265 tracking camera for localization and an RGB camera with a FoV of 120 degrees for observation generation. Each BMAV is equipped with an IMU

and downward-facing optical flow sensor. Meanwhile, each BMAV mounts a $3cm \times 3cm$ AprilTag for recognition and observation generation [24] (Fig. 6b). The AMAV adopts ArduPilot frameworks for motion control (Fig. 6c). A motion capture system provides millimeter-level ground truth at 240 FPS in the experiment area of $8m \times 8m$ (Fig. 6a and Fig. 6d). TransformLoc runs on a server featuring 128GB of memory and an Intel(R) Xeon(R) Gold 6242R CPU. We test the robustness of TransformLoc on a physical-feature-based simulator in an experiment area similar to Fig 6a.

Experiment setting. The BMAVs have a maximum velocity of $0.5m/s$ and a command interval δ of 5 steps. The AMAVs have an observation angle θ of 120 degrees and maximum observation distance r_m of $1m$. The control commands of AMAVs use motion primitives $\{(u, \omega)\} | u \in \{0, 1, 3\}m/s, \omega \in \{0, \pm 1, \pm 3\}rad/s$. We evaluated the performance by conducting in-field experiments with two AMAVs and six BMAVs, and simulations with five AMAVs and twenty BMAVs. The BMAVs move toward the edge of the room for environmental sensing. The simulations are conducted for 420 seconds, consistent with the typical battery life of the DJI Robomaster TT utilized in in-field experiments. The standard deviation of noise models for BMAVs' motion, AMAVs' range and bearing measurements are determined based on in-field experiments and set to 20%, 10%, and 5% of the measured values, respectively.

Comparative Methods. We tested TransformLoc (TL) against three baseline methods that do not require a localization infrastructure. ① Dead-Reckoning (DR) estimates the location of BMAVs with only measurements from motion sensors [28]; ② Station (ST) utilizes AMAVs with fixed locations to generate observations for BMAVs [29]; ③ H-SwarmLoc (SL) [30] is a SOTA method that navigates one AMAV based on reinforcement learning to assist BMAVs in localization. To ensure a fair comparison, we standardize SL to group MAVs in a consistent manner and direct AMAV to generate observation for BMAVs within the same group during navigation [30].

Evaluation Metrics. TransformLoc aims to improve the localization accuracy and navigation success rate of BMAVs. We use two metrics: ① Localization error: we compares the localization error of different BMAVs at each timestep, which is known as *absolute trajectory error* (ATE); ② Success rate: this is measured by counting the ratio of BMAVs that reach destinations under constraints.

B. Overall Performance

In-field experiments. Fig. 7 specifically focuses on in-field experimental results. Regarding localization error, a 420-second random walk is conducted for BMAVs, and the cumulative distribution function (CDF) of the ATE is plotted. TransformLoc achieves an ATE below $1.5m$, while the SL remains below $2.4m$, the ST below $2.8m$, and the DR below $3m$ (Fig. 7a). Regarding navigation performance, the success rate of BMAVs over a 200-second duration increases for all methods as the destination accuracy decreases (Fig. 7b).

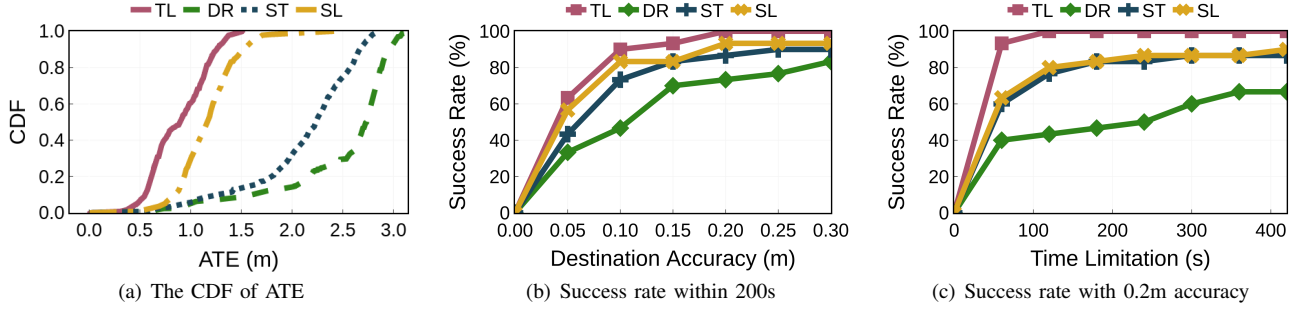


Fig. 7. Overall performance in in-field experiments with two AMAVs and six BMAVs. Our TransformLoc outperforms the baselines, demonstrating superior performance in both localization error and navigation success rate.

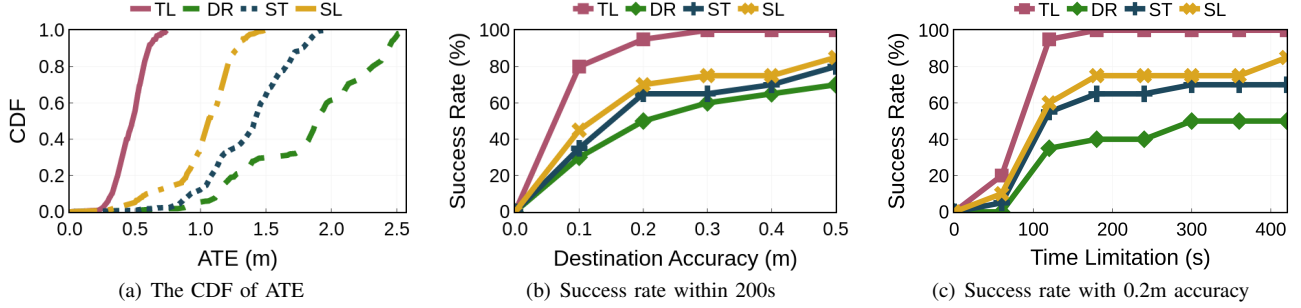


Fig. 8. Overall performance in physical feature-based simulation with five AMAVs and twenty BMAVs. Even with the increased numbers of AMAVs and BMAVs, TransformLoc continued to outperform the baseline in terms of localization accuracy and navigation success rate.

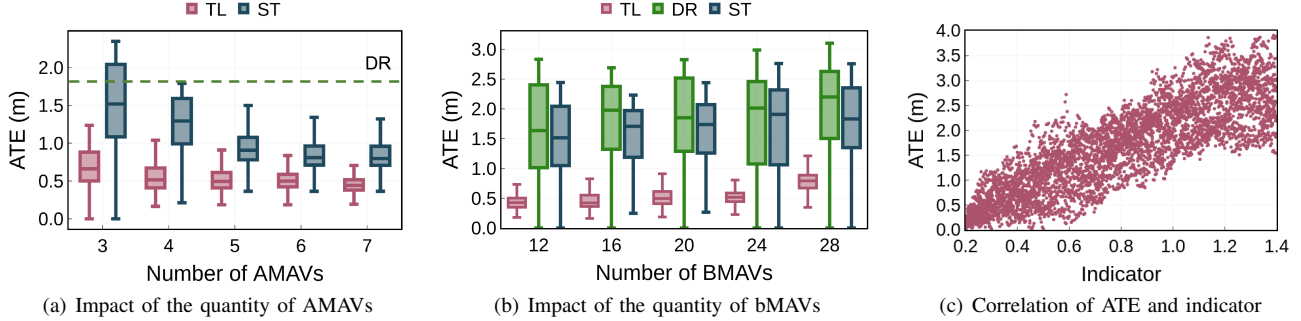


Fig. 9. System Robustness Evaluation with five AMAVs and twenty BMAVs. The experimental results indicate that TransformLoc is capable of adapting to changes in the quantities of AMAVs and BMAVs. Moreover, our indicator can infer the localization errors of BMAVs.

TransformLoc surpasses all baselines, achieving a remarkable 63% success rate with a stringent destination accuracy constraint of 0.05m. Furthermore, under looser destination accuracy limits (0.2m-0.3m), TransformLoc attains a 100% navigation success rate, outperforming the baselines. Additionally, as the time limitation increases, the success rate of BMAVs with a destination accuracy constraint of 0.2m improves for all methods, as they have more time to reach the target (Fig. 7c). TransformLoc exhibits superior performance compared to the baselines, achieving a navigation success rate 22.3% higher than SL, 25.6% higher than ST, and 55.6% higher than DR within strict time constraints (60-180 seconds). The above results illustrate enhancements in performance achieved by TransformLoc as it efficiently allocates AMAVs' resources to bolster the localization and navigation capabilities of BMAVs.

Physical feature-based simulation. Fig. 8 illustrates the performance of TransformLoc in the physical-feature-based simulation. In terms of localization error, TransformLoc

achieves an impressive ATE below 0.7m, while SL, ST, and DR maintain ATE below 1.5m, 2m, and 2.5m respectively (Fig. 8a). Regarding navigation success rate, TransformLoc consistently outperforms SL, ST, and DR by significant margins of 25%, 26.67%, and 33.3% on average, respectively, with varying destination accuracy over a 200-second duration (Fig. 8b). Furthermore, under a destination accuracy of 0.2m, TransformLoc surpasses SL, ST, and DR by an average of 22.8%, 25.6%, and 42.5% respectively, with varying time limitations (Fig. 8c). These results highlight the remarkable performance improvements achieved by TransformLoc.

C. System Robustness Evaluation

Impact of Number of AMAVs. Fig. 9a compares ATE of BMAVs against varying numbers of AMAVs with different methods. The TransformLoc exhibits lower errors than DR, which means the introduction of AMAVs reduced ATE. With the number of AMAVs increasing from 1 to 5, ATE of TransformLoc decreases below 0.8m, significantly lower

than ST which only helped BMAVs when passing over fixed AMAVs. Overall, the results indicate that increasing the number of AMAVs significantly reduces ATE of BMAVs.

Impact of Number of BMAVs. Fig. 9b shows how different methods perform when localizing varying numbers of BMAVs. As the number of BMAVs increases from 12 to 28, the ATE of all methods increases due to fewer percentage of BMAVs passing through AMAVs. However, TransformLoc consistently achieves an ATE of less than $1.2m$, which is more than 57% less than any other baselines. These experimental results show that TransformLoc is able to allocate sensing resources of AMAVs to provide observations for different numbers of BMAVs, resulting in a low error.

Effectiveness of indicator. Fig. 9c shows the effectiveness of our indicator. The indicator ($tr(\Sigma_{j,t})$) grows with ATE of BMAV, indicating that this indicator reflects the extent of localization error.

VI. RELATED WORK

Multi-agent for environmental sensing. Advancements in AI have fueled extensive research into multi-agent systems, harnessing their parallel operational pipelines and coordinated complementarity to improve sensing coverage and reduce time requirements for sensing tasks [5]. Agents in these systems can integrate various sensors, such as cameras [31], Radar [22], Lidar [32], IMU [20], acoustic sensors [19], and gas sensors [33]. This versatility enables the execution of diverse sensing tasks, including urban monitoring [34], [35], hazardous gas sourcing [36], and post-disaster data communication [37]. The data collected by these agents is shared through communication channels, providing insights into the agents' tasks and individual states, including motion policies [38], [39].

Localization and navigation of MAV swarm. In the realm of sensing tasks, precise localization and navigation play a pivotal role in facilitating effective collaboration among MAV swarms, particularly given the inherent constraints in computing, communication, and sensing capabilities of individual MAVs [40]. Simultaneous Localization and Mapping (SLAM) stands out as the most widely employed method [41]. This method entails outfitting MAVs with a suite of sensors, including RGB cameras, LiDAR, and depth cameras, to gather information about the environment and their own states. Subsequently, navigation algorithms are deployed to guide MAVs to their intended destinations [15], [42]. While collaborative SLAM utilizing multiple MAVs has been investigated in prior studies [31], the prohibitive cost of sensors imposes limitations on the scalability of MAV swarms [11]. In response to this challenge, approaches centered on external infrastructure have been proposed. For instance, radio frequency-based localization offers high accuracy at a relatively low cost of sensors [14]. However, these approaches rely on the presence of additional installed localization infrastructure, presenting challenges in environments where localization infrastructure may have been destroyed

or where such installations are impractical, especially in hazardous conditions [43].

To address these constraints, we propose a collaborative and adaptive localization framework, named *TransformLoc*, for a heterogeneous MAV swarm. Within this framework, a group of AMAVs serves as mobile localization infrastructure, collaboratively sharing sensing and computing capabilities with a larger number of BMAVs. This cooperative approach enables precise localization for a MAV swarm at a more economical cost.

VII. DISCUSSION

We delve into several influential factors of TransformLoc.

- **Communication Load:** The BMAV transmits location estimates (mean and covariance) to AMAVs, and AMAV transmits motion commands to BMAVs. When catering to twenty BMAVs, an AMAV both receives and transmits less than 10KB of data every δ seconds. During grouping, an AMAV gathers location information from others and transmits grouping results to others. with twenty AMAVs, the data volume stays below 10KB. In summary, the communication load is deemed manageable.
- **Localization accuracy of AMAV:** The system improves the localization accuracy of BMAVs when the localization error of AMAVs is less than 10cm. Achieving this level of accuracy is easily feasible by outfitting an AMAV with a Lidar or depth camera and employing SLAM methods, such as VINS (Visual-Inertial Navigation System).

VIII. CONCLUSION

In this paper, we propose the design of TransformLoc, a new framework that transforms AMAVs into mobile localization infrastructures. The innovation of TransformLoc lies in two aspects: 1) we derive an error-aware joint location estimation model to integrate inaccurate estimation from BMAVs with observations from AMAVs to perform joint location estimation of BMAVs, assisted by an uncertainty-aided inference method; and 2) we design a proximity-driven adaptive grouping-scheduling strategy to dynamically allocate AMAVs' sensing resources to assist BMAVs. The evaluation through in-field experiments and large-scale physical feature-based simulations demonstrate the superior performance of the TransformLoc.

IX. ACKNOWLEDGMENTS

This paper was supported by the National Key R&D program of China No. 2022YFC3300703, the Natural Science Foundation of China under Grant No. 62371269. Guangdong Innovative and Entrepreneurial Research Team Program No. 2021ZT09L197, Shenzhen 2022 Stabilization Support Program No. WDZC20220811103500001, and Tsinghua Shenzhen International Graduate School Cross-disciplinary Research and Innovation Fund Research Plan No. JC20220011. We acknowledge the support from the Tsinghua Shenzhen International Graduate School-Shenzhen Pengrui Endowed Professorship Scheme of Shenzhen Pengrui Foundation.

APPENDIX A KEY DEFINITIONS

A. Environmental Description

Let Ω be a bounded space in $\mathbb{R}^3$ with length, width, and height L , W , and H , respectively, where the heterogeneous MAV swarm operates. This swarm comprises a fixed number of BMAVs and AMAVs that operate independently. For simplicity, we assume that both types of MAVs operate at the same altitude in this paper.

B. State model of AMAV

The heterogeneous MAV swarm contains M ($M > 1$) AMAVs. The state of each AMAV A_j at time t consists of the location and the motion command. The location is denoted as $\mathbf{x}_{j,t} = (x_{j,t}^1, x_{j,t}^2, \phi_{j,t})$, where $\phi_{j,t}$ is the orient angle. The distance between two locations is d_χ . For motion command, $\mathbf{u}_{j,t} = (u_{j,t}, \omega_{j,t})$, where $u_{j,t}$ and $\omega_{j,t}$ are the translational and rotational velocities, respectively. The AMAVs estimate their location accurately with advanced sensing capabilities and follow the motion model $\mathbf{x}_{j,t} = f(\mathbf{x}_{j,t-1}, \mathbf{u}_{j,t-1})$, which can be represented as follows:

$$\begin{pmatrix} x_{j,t}^1 \\ x_{j,t}^2 \\ \phi_{j,t} \end{pmatrix} = \begin{pmatrix} x_{j,t-1}^1 \\ x_{j,t-1}^2 \\ \phi_{j,t-1} \end{pmatrix} + \begin{pmatrix} u_{j,t-1} \cos(\phi_{j,t-1}) \\ u_{j,t-1} \sin(\phi_{j,t-1}) \\ \omega_{j,t-1} \end{pmatrix}. \quad (3)$$

C. State model of BMAV

The heterogeneous MAV swarm includes N ($N > 1$) BMAVs. The state of each BMAV B_i at time t consists of the location and the motion command. The location of B_i is $\mathbf{y}_{i,t} = (y_{i,t}^1, y_{i,t}^2)$, and its motion command is $\mathbf{v}_{i,t} = (v_{i,t}^1, v_{i,t}^2)$. B_i follows double integrator dynamics with Gaussian noise represented as follows:

$$\begin{pmatrix} y_{i,t+\delta}^1 \\ y_{i,t+\delta}^2 \end{pmatrix} = \begin{pmatrix} y_{i,t}^1 \\ y_{i,t}^2 \end{pmatrix} + \delta \begin{pmatrix} v_{i,t}^1 + n_{i,t}^1 \\ v_{i,t}^2 + n_{i,t}^2 \end{pmatrix}, \quad (4)$$

$n_{i,t}^1$ is drawn from $p(n^1)$,
 $n_{i,t}^2$ is drawn from $p(n^2)$,

where $n_{i,t}^1$ and $n_{i,t}^2$ are drawn from motion noise models $p(n^1)$ and $p(n^2)$ respectively, δ ($\delta > 1$) denotes the time interval between two commands. $p(n^1)$ and $p(n^2)$ empirically obtained from the testbed are specified as normal distributions with mean $\mu = 0$ and variance σ , expressed as a percentage of $v_{i,t}^1$ and $v_{i,t}^2$, same as outlined in [44]. The location estimation of B_i is denoted as $\hat{\mathbf{y}}_{i,t}$, and the covariance matrix is denoted as $\Sigma_{i,t}$. The noisy motion measurement from B_i 's IMU is denoted as $\hat{\mathbf{v}}_{i,t}$.

D. Observation model of AMAV

The observation angle of the visual sensor is denoted by ϕ , and the FoV of A_j at time t is defined as follows,

$$F_{j,t} = \{(x, y) \mid 0 < \sqrt{(y - x_{j,t}^2)^2 + (x - x_{j,t}^1)^2} < r_m \cap -\frac{\phi}{2} < \arctan((y - x_{j,t}^2)(x - x_{j,t}^1)) - \phi < \frac{\phi}{2}\}, \quad (5)$$

where r_m is the maximum distance that the sensor can observe, which is limited by the observation technique.

When B_i in the FoV of A_j at time t , A_j generates noisy observation $z_{i,j,t}$ for B_i , which consists of range $r_{i,j,t}$ and bearing $\alpha_{i,j,t}$ for B_i , both $r_{i,j,t}$ and $\alpha_{i,j,t}$ are relative to A_i ,

$$\begin{aligned} z_{i,j,t} &= h(\mathbf{x}_{j,t}, \mathbf{y}_{i,t}) + \mathbf{n}_{i,j,t}, \\ h(\mathbf{x}_{j,t}, \mathbf{y}_{i,t}) &:= \begin{bmatrix} r_{i,j,t} \\ \alpha_{i,j,t} \end{bmatrix} \\ &:= \begin{bmatrix} \sqrt{(y_{i,t}^1 - x_{j,t}^1)^2 + (y_{i,t}^2 - x_{j,t}^2)^2} \\ \arctan((y_{i,t}^2 - x_{j,t}^2)(y_{i,t}^1 - x_{j,t}^1)) - \phi \end{bmatrix}, \\ \mathbf{n}_{i,j,t} &:= [n_{i,j,t}^r, n_{i,j,t}^\alpha]^T \\ n_{i,j,t}^r &\text{ is drawn from } p(n^r), \\ n_{i,j,t}^\alpha &\text{ is drawn from } p(n^\alpha). \end{aligned} \quad (6)$$

The noise models for range and bearing measurements are denoted as $p(n^r)$ and $p(n^\alpha)$, respectively, and are empirically obtained from our testbed. $p(n^r)$ and $p(n^\alpha)$ are specified as a normal distribution with mean $\mu = 0$ and variance σ , expressed as a percentage of $r_{i,j,t}$ and $\alpha_{i,j,t}$.

E. Linearization of Observation model

We compute the Jacobian matrix of $h(\mathbf{x}, \mathbf{y})$ based on Taylor expansion by taking the gradient with respect to the location estimation of $\mathbf{y}$ to linearize the observation model,

$$\begin{aligned} \nabla_{\mathbf{y}} h(\mathbf{x}, \mathbf{y}) &= \frac{1}{r(\mathbf{x}, \mathbf{y})} \begin{bmatrix} (y^1 - x^1) & (y^2 - x^2) \\ -\sin(\phi + \alpha(\mathbf{x}, \mathbf{y})) & \cos(\phi + \alpha(\mathbf{x}, \mathbf{y})) \end{bmatrix}. \end{aligned} \quad (7)$$

To simplify notation, we let $\mathbf{y}_t := [\mathbf{y}_{1,t}^T, \dots, \mathbf{y}_{N,t}^T]^T$, $\mathbf{x}_t := [\mathbf{x}_{1,t}^T, \dots, \mathbf{x}_{M,t}^T]^T$, $\mathbf{u}_t := [\mathbf{u}_{1,t}^T, \dots, \mathbf{u}_{M,t}^T]^T$, $\hat{\mathbf{y}}_t := [\hat{\mathbf{y}}_{1,t}^T, \dots, \hat{\mathbf{y}}_{N,t}^T]^T$, $\Sigma_t := \text{diag}(\Sigma_{1,t}^T, \dots, \Sigma_{N,t}^T)^T$.

APPENDIX B PROBLEM FORMULATION

We aim to optimize the trace of the covariance of BMAVs' location estimation Σ_t to limit the localization error within a finite time horizon T , considering the motion commands of each AMAV within T . The mathematically formulated sensing resource allocation problem is given by,

$$\begin{aligned} \min_{\mathbf{u}_0, \dots, \mathbf{u}_{T-1}} \quad & \xi(T) = \sum_{t=1}^T (\text{tr}(\Sigma_t)), \\ \text{s.t.} \end{aligned} \quad (8)$$

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t), t \in \{0, \dots, T-1\} \quad (9)$$

$$0 \leq x_{j,t}^1, y_{i,t}^1 \leq L, \quad (10)$$

$$j \in \{0, \dots, N\}, i \in \{0, \dots, M\}, t \in \{0, \dots, T-1\}$$

$$0 \leq x_{j,t}^2, y_{i,t}^2 \leq W, \quad (11)$$

$$j \in \{0, \dots, N\}, i \in \{0, \dots, M\}, t \in \{0, \dots, T-1\}$$

$$\Sigma_{t+1} = \rho_{t+1}^e (\rho_t^p(\Sigma_t), \mathbf{x}_{t+1}), t \in \{0, \dots, T-1\}. \quad (12)$$

ρ_t^e and ρ_{t+1}^p are correction and prediction steps, which are described in Section III-B.

REFERENCES

- [1] M. T. Rashid, D. Y. Zhang, and D. Wang, "Socialdrone: An integrated social media and drone sensing system for reliable disaster response," in *Proceedings of the IEEE INFOCOM*, 2020, pp. 218–227.
- [2] K. McGuire, C. De Wagter, K. Tuyls, H. Kappen, and G. C. de Croon, "Minimal navigation solution for a swarm of tiny flying robots to explore an unknown environment," *Science Robotics*, vol. 4, no. 35, p. eaaw9710, 2019.
- [3] A. Khochare, Y. Simmhan, F. B. Sorbelli, and S. K. Das, "Heuristic algorithms for co-scheduling of edge analytics and routes for uav fleet missions," in *Proceedings of the IEEE INFOCOM*, 2021, pp. 1–10.
- [4] L. Bertizzolo, S. D'oro, L. Ferranti, L. Bonati, E. Demirors, Z. Guan, T. Melodia, and S. Pudlewski, "Swarmcontrol: An automated distributed control framework for self-optimizing drone networks," in *Proceedings of the IEEE INFOCOM*, 2021, pp. 1768–1777.
- [5] Y. Wang, Z. Su, Q. Xu, R. Li, and T. H. Luan, "Lifesaving with rescuechain: Energy-efficient and partition-tolerant blockchain based secure information sharing for uav-aided disaster rescue," in *Proceedings of the IEEE INFOCOM*, 2021, pp. 1–10.
- [6] J. Li, H. Kang, G. Sun, S. Liang, Y. Liu, and Y. Zhang, "Physical layer secure communications based on collaborative beamforming for uav networks: A multi-objective optimization approach," in *Proceedings of the IEEE INFOCOM*, 2021, pp. 1–10.
- [7] "Global drones market outlook (2022-2032)," <https://www.factmr.com/report/62/drone-market>.
- [8] C. Ruiz, X. Chen, L. Zhang, and P. Zhang, "Collaborative localization and navigation in heterogeneous uav swarms: Demo abstract," in *Proceedings of the 14th ACM Sensys*, 2016, pp. 324–325.
- [9] X. Chen, A. Purohit, S. Pan, C. Ruiz, J. Han, Z. Sun, F. Mokaya, P. Tague, and P. Zhang, "Design experiences in minimalistic flying sensor node platform through sensorfly," *ACM TOSN*, vol. 13, no. 4, pp. 1–37, 2017.
- [10] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The euroc micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, 2016.
- [11] P. Schmuck and M. Chli, "Ccm-slam: Robust and efficient centralized collaborative monocular simultaneous localization and mapping for robotic teams," *Journal of Field Robotics*, vol. 36, no. 4, 2019.
- [12] T. Li, S. Leng, Z. Wang, K. Zhang, and L. Zhou, "Intelligent resource allocation schemes for uav-swarm-based cooperative sensing," *IEEE Internet of Things Journal*, vol. 9, no. 21, pp. 21 570–21 582, 2022.
- [13] A. Trotta, F. D. Andreagiovanni, M. Di Felice, E. Natalizio, and K. R. Chowdhury, "When uavs ride a bus: Towards energy-efficient city-scale video surveillance," in *Proceedings of the IEEE INFOCOM*, 2018, pp. 1043–1051.
- [14] G. Chi, Z. Yang, J. Xu, C. Wu, J. Zhang, J. Liang, and Y. Liu, "Wi-drone: Wi-fi-based 6-dof tracking for indoor drone flight control," in *Proceedings of the ACM MobiSys*, 2022.
- [15] J. Xu, H. Cao, D. Li, K. Huang, C. Qian, L. Shangguan, and Z. Yang, "Edge assisted mobile semantic visual slam," in *Proceedings of the IEEE INFOCOM*, April 27–30 2020.
- [16] J. Sharp, C. Wu, and Q. Zeng, "Authentication for drone delivery through a novel way of using face biometrics," in *Proceedings of the 28th ACM MobiCom*, 2022, pp. 609–622.
- [17] S. Xu, X. Chen, X. Pi, C. Joe-Wong, P. Zhang, and H. Y. Noh, "ilocus: Incentivizing vehicle mobility to optimize sensing distribution in crowd sensing," *IEEE Transactions on Mobile Computing*, vol. 19, no. 8, pp. 1831–1847, 2019.
- [18] D. Li, J. Xu, Z. Yang, Y. Lu, Q. Zhang, and X. Zhang, "Train once, locate anytime for anyone: Adversarial learning based wireless localization," in *Proceedings of the IEEE INFOCOM*, May 10–13 2021.
- [19] W. Wang, L. Mottola, Y. He, J. Li, Y. Sun, S. Li, H. Jing, and Y. Wang, "Micnest: Long-range instant acoustic localization of drones in precise landing," in *Proceedings of the 20th ACM SenSys*, 2022.
- [20] X. Chen, A. Purohit, C. R. Dominguez, S. Carpin, and P. Zhang, "Drunkwalk: Collaborative and adaptive planning for navigation of micro-aerial sensor swarms," in *Proceedings of the 13th ACM Sensys*, 2015, pp. 295–308.
- [21] Y. Sun, W. Wang, L. Mottola, R. Wang, and Y. He, "Aim: Acoustic inertial measurement for indoor drone localization and tracking," in *Proceedings of the 20th ACM Sensys*, 2022, pp. 476–488.
- [22] C. X. Lu, M. R. U. Saputra, P. Zhao, Y. Almalioğlu, P. P. De Gusmao, C. Chen, K. Sun, N. Trigoni, and A. Markham, "milliego: single-chip mmwave radar aided egomotion estimation via deep sensor fusion," in *Proceedings of the 18th Sensys*, 2020, pp. 109–122.
- [23] E. Dong, J. Xu, C. Wu, Y. Liu, and Z. Yang, "Pair-navi: Peer-to-peer indoor navigation with mobile visual slam," in *Proceedings of the IEEE INFOCOM*, April 29–May 2 2019.
- [24] J. Wang and E. Olson, "Apritag 2: Efficient and robust fiducial detection," in *Proceedings of the IEEE IROS*, 2016, pp. 4193–4198.
- [25] J. Xu, G. Chi, Z. Yang, D. Li, Q. Zhang, Q. Ma, and X. Miao, "Followupar: Enabling follow-up effects in mobile ar applications," in *Proceedings of the ACM MobiSys*, June 24–July 2 2021.
- [26] D. Ucinski, *Optimal measurement methods for distributed parameter system identification*. CRC press, 2004.
- [27] X. Zhang, A. Zhang, J. Sun, X. Zhu, Y. E. Guo, F. Qian, and Z. M. Mao, "Emp: Edge-assisted multi-vehicle perception," in *Proceedings of the 27th ACM Mobicom*, 2021, pp. 545–558.
- [28] J. Borenstein, H. Everett, and L. Feng, *Navigating mobile robots: Systems and techniques*. AK Peters, Ltd., 1996.
- [29] J. Xu, F. Zhong, and Y. Wang, "Learning multi-agent coordination for enhancing target coverage in directional sensor networks," *Proceedings of the NeurIPS*, vol. 33, pp. 10 053–10 064, 2020.
- [30] H. Wang, X. Chen, Y. Cheng, C. Wu, F. Dang, and X. Chen, "H-swarmloc: Efficient scheduling for localization of heterogeneous mav swarm with deep reinforcement learning," in *Proceedings of the 20th ACM Sensys*, 2022, pp. 1148–1154.
- [31] J. Xu, H. Cao, Z. Yang, L. Shangguan, J. Zhang, X. He, and Y. Liu, "Swarmmap: Scaling up real-time collaborative visual slam at the edge," in *Proceedings of the USENIX NSDI*, 2022, pp. 977–993.
- [32] D. Li, J. Xu, Z. Yang, Q. Zhang, Q. Ma, L. Zhang, and P. Chen, "Motion inspires notion: self-supervised visual-lidar fusion for environment depth estimation," in *Proceedings of the 20th MobiSys*, 2022, pp. 114–127.
- [33] J. Luo, Y. Hu, C. Yu, C. Hong, X.-P. Zhang, and X. Chen, "Field reconstruction-based non- rendezvous calibration for low cost mobile sensors," in *Proceedings of the ACM Ubicomp*, 2023, pp. 688–693.
- [34] J. Guo, H. Wang, W. Liu, G. Huang, J. Gui, and S. Zhang, "A lightweight verifiable trust based data collection approach for sensor-cloud systems," *Journal of Systems Architecture*, vol. 119, p. 102219, 2021.
- [35] X. Chen, H. Wang, Z. Li, W. Ding, F. Dang, C. Wu, and X. Chen, "Deliversense: Efficient delivery drone scheduling for crowdsensing with deep reinforcement learning," in *Proceedings of the ACM Ubicomp*, 2022, pp. 403–408.
- [36] Y. Liu, X. Liu, F. Man, C. Wu, and X. Chen, "Fine-grained air pollution data enables smart living and efficient management," in *Proceedings of the 20th ACM Sensys*, 2022, pp. 768–769.
- [37] J. Ren, Y. Xu, Z. Li, C. Hong, X.-P. Zhang, and X. Chen, "Scheduling uav swarm with attention-based graph reinforcement learning for ground-to-air heterogeneous data communication," in *Proceedings of the ACM Ubicomp*, 2023, pp. 670–675.
- [38] X. Chen, S. Xu, X. Liu, X. Xu, H. Y. Noh, L. Zhang, and P. Zhang, "Adaptive hybrid model-enabled sensing system (hmss) for mobile fine-grained air pollution estimation," *IEEE Transactions on Mobile Computing*, vol. 21, no. 6, pp. 1927–1944, 2020.
- [39] L. Zhou, S. Leng, Q. Liu, and Q. Wang, "Intelligent uav swarm cooperation for multiple targets tracking," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 743–754, 2021.
- [40] Z. Li, F. Man, X. Chen, B. Zhao, C. Wu, and X. Chen, "Tract: Towards large-scale crowdsensing with high-efficiency swarm path planning," in *Proceedings of the ACM Ubicomp*, 2022, pp. 409–414.
- [41] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.
- [42] X. Zhou, X. Wen, Z. Wang, Y. Gao, H. Li, Q. Wang, T. Yang, H. Lu, Y. Cao, C. Xu *et al.*, "Swarm of micro flying robots in the wild," *Science Robotics*, vol. 7, no. 66, p. eabm5954, 2022.
- [43] S. Kumar, S. Gil, D. Katabi, and D. Rus, "Accurate indoor localization with zero start-up cost," in *Proceedings of the ACM MobiCom*, 2014, pp. 483–494.
- [44] X. Chen, C. Ruiz, S. Zeng, L. Gao, A. Purohit, S. Carpin, and P. Zhang, "H-drunkwalk: Collaborative and adaptive navigation for heterogeneous mav swarm," *ACM Transactions on Sensor Networks*, vol. 16, no. 2, pp. 1–27, 2020.