2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

Swarm-LIO: Decentralized Swarm LiDAR-inertial Odometry

Fangcheng Zhu*, Yunfan Ren*, Fanze Kong, Huajie Wu, Siqi Liang, Nan Chen, Wei Xu, Fu Zhang

***Abstract*— Accurate self and relative state estimation are the critical preconditions for completing swarm tasks, *e.g.*, collaborative autonomous exploration, target tracking, search and rescue. This paper proposes Swarm-LIO: a fully decentralized state estimation method for aerial swarm systems, in which each drone performs precise ego-state estimation, exchanges ego-state and mutual observation information by wireless communication, and estimates relative state with respect to (w.r.t.) the rest of UAVs, all in real-time and only based on LiDAR-inertial measurements. A novel 3D LiDAR-based drone detection, identification and tracking method is proposed to obtain observations of teammate drones. The mutual observation measurements are then tightly-coupled with IMU and LiDAR measurements to perform real-time and accurate estimation of ego-state and relative state jointly. Extensive real-world experiments show the broad adaptability to complicated scenarios, including GPS-denied scenes, degenerate scenes for camera (dark night) or LiDAR (facing a single wall). Compared with ground-truth provided by motion capture system, the result shows the centimeter-level localization accuracy which outperforms other state-of-the-art LiDAR-inertial odometry for single UAV system.**

I. Introduction

Multi-robot system, especially aerial swarm system, has great potential in many aspects, such as autonomous exploration [1], [2], target tracking [3], search and rescue, etc. Thanks to its great team cooperation capability, a swarm system can complete missions in complex scenarios even in degenerate environment for a single drone. For a single drone system, accurate self-localization [4]–[7] is the precondition for obstacle avoidance [8]–[10] and flight control [11]. For an aerial swarm, robust, accurate self and relative state estimation [12] play a similarly vital role. For a swarm to fulfill cooperative tasks, each drone in the swarm needs to perform real-time, precise self-localization with low drift, and needs to be aware of other drones' states at all times.

Most recent works on aerial swarm mainly focus on collaborative planning [13]–[15], while research on self and relative state estimation of swarm system is still a large gap. For outdoor scenes, GPS and RTK-GPS are usually adopted [16], [17]. While for GPS-denied places such as indoor scenarios, motion capture system [15], visual (-inertial) odometry [14], [18] or LiDAR (-inertial) odometry [1], [19] methods are much preferred. Besides, Ultra-WideBand (UWB) is also adopted in [13], [20] to produce robust localization result. Although RTK-GPS, motion capture and UWB with anchors [21] have great accuracy, they are dependent on cumbersome, extra devices and would lead to laborious installation. Moreover, these mentioned methods render the whole system centralized, which is fragile to single-point-of-failure (SPOF). Camera is widely used due to its light weight, low cost and rich color information, but it is vulnerable to inadequate illumination and is lack direct depth measurement, leading to high computational complexity when computing 3D measurements. Apart from that, the range of depth measurements of camera is quite limited. Anchor-free UWB is also low-cost and light-weight. However, it can only offer one-dimensional distance measurement and the accuracy is quite limited (usually meter-level), which may decrease the overall accuracy of the whole swarm system.

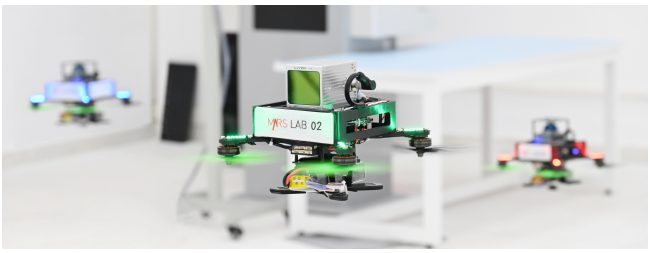

Fig. 1. Indoor flights of the proposed aerial swarm with three drones.

Compared with camera and anchor-free UWB, LiDAR can provide accurate 3D measurements and is also robust to illumination changes. Even in dark scenes, *e.g.* underground tunnels, LiDAR can help to perform accurate localization and mapping [18]. In recent years, some cost-effective solid-state LiDARs such as Livox[1] emerged in the market, which significantly expanded the scope of LiDAR-based research. In this paper, we propose a robust, real-time and decentralized swarm odometry based on LiDAR-inertial measurements. The main contributions are highlighted as follows:

1) A novel 3D LiDAR-based drone detection, identification and tracking method, providing accurate 3D mutual observation measurements used for self and relative state estimation. Each drone is attached with reflective tapes, such that a teammate drone can be reliably and efficiently detected from the reflectivity values of LiDAR points. The detected teammate drone is then tracked by a Kalman filter and matched with trajectories received from the shared network to obtain the teammate's identity and its initial relative state.
2) A fully decentralized, robust, data-effective and computationally-efficient ego-state and relative state estimation framework. In this framework, each drone only has to exchange its ego-state and teammate observation information, which requires extremely low communica-

*These two authors contribute equally to this work.

F. Zhu, Y. Ren, F. Kong, H. Wu and F. Zhang are with Department of Mechanical Engineering, The University of Hong Kong. {*zhufc,renyf,kongfz,wu2020,cnchen*}*@connect.hku.hk*, {*xuweii,fuzhang*}*@hku.hk*. S. Liang is with Harbin Institute of Technology, Shenzhen *sqliang@stu.hit.edu.cn*

[1]https://www.livoxtech.com

2023 IEEE International Conference on Robotics and Automation (ICRA) | 979-8-3503-2365-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICRA48891.2023.10161355

Authorized licensed use limited to: ShanghaiTech University. Downloaded on October 05,2024 at 02:37:33 UTC from IEEE Xplore. Restrictions apply.

tion bandwidth. The mutual observation measurement is tightly-coupled with LiDAR and IMU measurements under Error State Iterated Kalman Filter (ESIKF) [22] framework to achieve robust, accurate, and simultaneous self and relative state estimation.

3) Low-cost, decentralized hardware and software framework. The sensors and computing resources are totally on-board, including Livox LiDAR, IMU and on-board computer. The communication is achieved with Ad-Hoc network available on standard network module.
4) Implementation and validation of the proposed method in extensive real-world experiments, in both indoor (see Fig. 1), outdoor and degenerate scenes.

II. Related Works

State estimation for aerial swarm systems is much more complicated than that of single drone system because each drone not only needs to perform accurate self-localization with low drift, but also needs to estimate teammates' states in its own global frame. Among these swarm systems, cameras and LiDARs are two mainstream sensors.

Many existing works on swarm state estimation are based on image measurements. In [14], [23] each drone in the swarm runs an independent VIO to achieve ego-state estimation. The ego-state is then exchanged so that each drone can estimate the relative state of other teammates given the relative transformation between their initial pose (global extrinsic). The global extrinsic is usually calibrated offline [14] or estimated online with good initial guess [12]. Since the VIO of each drone is completely independent, it may lead to large relative state estimation error as the system runs, due to imprecise global extrinsic or VIO drift caused by long-time running. To constrain the relative state during the online running of swarm systems, mutual observation based methods are vigorously developing. Nguyen *et al*. [24] propose visual-inertial multi-drone localization system, MAVNet is used to detect other teammate drones. To further enhance the mutual observation, Xu *et al*. [12] propose a visual-inertial-UWB relative state estimation system utilizing YOLOv3-tiny for teammate drone detection, and UWB module is adopted to provide distance constraints as a complementary sensor of camera. These learning-based drone detection methods usually need a long time of prior training, resulting in increment of the total time-consumption of system implementation. One of the biggest restrictions of vision-based methods is the range and quality of depth measurement are quite limited, which narrows down their applications in practical cases, *e.g*. in large-scale, outdoor environments.

By contrast, LiDAR can provide accurate and long-range depth measurements, bringing many new opportunities for swarm state estimation. In [19], [21], [25], 3D LiDAR-based place recognition (loop closure) constraints are utilized to improve state estimation accuracy. However, since cloud register of raw or segmented points is inevitable in place recognition, the computation load is quite heavy. Thus, all [19], [21], [25] need a centralized master agent to receive data transmitted by other agents and finish those computationally intensive tasks, *i.e*., cloud register or map merging, increasing the risk of single-point-of-failure. To ease the burden of communication and computation, Huang *et al*. [26] propose a decentralized multi-robot system using Scan Context descriptor, permitting data-efficient exchange of LiDAR descriptors among robots. Apart from map and descriptor exchange mentioned above, mutual observation is another important unique feature of swarm system, which is a light-weight type of information, avoiding large communication and computation load. Wasik *et al*. [27] propose a laser-based multi-robot system, laser range finders are used for each robot to estimate the distances and angles to other robots. However, the used 2D LiDARs are not applicable to drones considered in this paper which fly in 3D spaces.

Compared with learning-based detection methods [12], [24], we attach reflective tapes on each drone and leverage the reflectivity measured by LiDAR sensors to detect teammate drones. Different from [12], [14], we propose a coarse-to-fine calibration method to acquire accurate global extrinsic transformations without any prior initial guess. The rough calibration result acquired by trajectory matching is fed into ESIKF for further online refinement along with ego-state estimation. Compared with centralized systems in which the drones exchange map information [19], [21], [25], our system is fully decentralized which would suffer no SPOF issue, and communication-efficient which exchanges ego-state and mutual observation information only.

III. Methodology

A. Problem Formulation

Aiming to assist understanding of the proposed system, we define some important notations here. $\mathbf{x}_i, \widehat{\mathbf{x}}_i, \bar{\mathbf{x}}_i, \breve{\mathbf{x}}_i$ represent the ground-truth, predictive, updated, and measured state of drone i respectively. t_{ik} denotes the timestamp of the k-th ESIKF state update of drone i.

Considering an aerial swarm system consisting of N drones and each one carries a LiDAR and an inertial measurement unit (IMU), the state estimation is decomposed into two parts. The first part is ego-state estimation. Each drone, here we choose drone i for demonstration convenience, needs to estimate its own position ${}^{G_i}\mathbf{p}_{b_i} \in \mathbb{R}^3$ and attitude ${}^{G_i}\mathbf{R}_{b_i} \in SO(3)$ in its global frame G_i. Since the IMU measurements are coupled with unknown and time-varying bias, the gyroscope bias $\mathbf{b}_{g_i} \in \mathbb{R}^3$ and accelerometer bias $\mathbf{b}_{a_i} \in \mathbb{R}^3$ should be calibrated. Moreover, gravity vector in each global frame ${}^{G_i}\mathbf{g} \in \mathbb{R}^3$ should also be estimated online. The self-position ${}^{G_i}\mathbf{p}_{b_i}$, attitude ${}^{G_i}\mathbf{R}_{b_i}$, velocity ${}^{G_i}\mathbf{v}_{b_i}$, bias $\mathbf{b}_{g_i}, \mathbf{b}_{a_i}$, and gravity vector ${}^{G_i}\mathbf{g}$ constitute the ego-state, whose estimation is exchanged among the swarm system. Then, to estimate the relative state of other teammate drones based on the exchanged ego-state information, each drone needs to estimate the global extrinsic transformation ${}^{G_i}\mathbf{T}_{G_j} = ({}^{G_i}\mathbf{R}_{G_j}, {}^{G_i}\mathbf{p}_{G_j}) \in SE(3)$ with respect to the teammate drone j ($j = 1, \cdots, N, j \neq i$). Finally, to provide mutual observations of other teammates, each drone i needs to detect, identify, and track any other teammate j, whose observed position in the self body frame is denoted by ${}^{b_i}\breve{\mathbf{p}}_{b_j}$.

Authorized licensed use limited to: ShanghaiTech University. Downloaded on October 05,2024 at 02:37:33 UTC from IEEE Xplore. Restrictions apply.

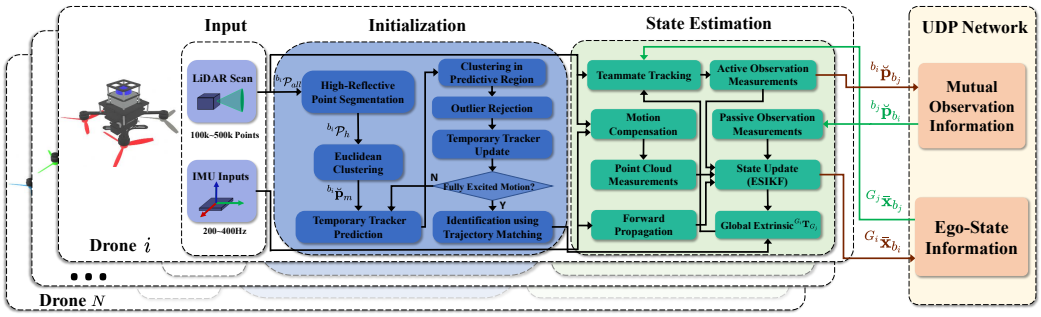

Fig. 2. Framework of the proposed state estimation system for aerial swarm systems.

B. Framework Overview

Each drone in the swarm system runs an independent copy of the same framework. The overview of the framework copy run by the i-th drone is shown Fig. 2. The first stage is initialization (Sec. III-C), including detection and temporary tracking for potential teammate drones (Sec. III-C.1). Once the motion of temporarily tracked object is sufficiently excited, the trajectory matching (Sec. III-C.2) starts. If the trajectory of the temporarily tracked object is successfully matched with a teammate's trajectory transmitted by wireless communication, the object would be considered as the observation of the corresponding teammate and be labeled with the teammate identity (ID), the temporary tracker hence becomes a teammate tracker (Sec. III-D.1). The observed position of teammate j, ${}^{b_i}\breve{\mathbf{p}}_{b_j}$ (the "active observation measurement"), the self-position observed by teammate j, ${}^{b_j}\breve{\mathbf{p}}_{b_i}$ (the "passive observation measurement", which is received from the network), the LiDAR point cloud (after motion compensation), and the IMU measurements are then fused by an error-state iterative Kalman filter (ESIKF) (Sec. III-D) to jointly estimate the ego-state and global extrinsic transformation ${}^{G_i}\mathbf{T}_{G_j}$. Finally, the estimated ego-state and active mutual observation information are transmitted to other drones via UDP packets via Ad-Hoc communication.

C. Initialization of Swarm System

In this section, we introduce how the drones detect, track and identify a potential teammate drone.

1) Drone Detection and Temporary Tracking: We propose a novel drone detection method based on reflectivity filtering and cluster extraction. For each drone, reflective tapes are attached to its body, so that it can be easily detected by other teammates based on the reflectivity information measured by LiDAR sensor. The detailed detection and tracking procedures at each LiDAR scan for drone i are summarized in Alg. 1. Denote ${}^{b_i}\mathcal{P}_{all}$ as the raw LiDAR points in a new scan represented in the current body frame. First, points with high reflectivity values exceeding a given threshold, which can be calibrated beforehand on the reflective tapes, are extracted by **ReflectivityFiltering**(${}^{b_i}\mathcal{P}_{all}$) in Line 1. Then in Line 2, the high-reflectivity points ${}^{b_i}\mathcal{P}_h$ are clustered by **Euclidean-Clustering**(${}^{b_i}\mathcal{P}_h$)[2] to detect new potential teammate drones. The newly detected object is then tracked by a Kalman filter-based temporary tracker in Line 4.

The state vector $\mathbf{x}_m$ of each temporary tracker m ($m = 1, 2, \cdots M$) consists of the object position ${}^{G_i}\mathbf{p}_m$ and velocity in drone i's global frame and the temporary trackers will predict a tracker position ${}^{G_i}\hat{\mathbf{p}}_m$ based on constant velocity model. Then, the position ${}^{G_i}\breve{\mathbf{p}}_m = {}^{G_i}\mathbf{T}_{b_i}\,{}^{b_i}\breve{\mathbf{p}}_m$ (where ${}^{b_i}\breve{\mathbf{p}}_m$ is clustered from high-reflectivity points in Line 2) is associated with the closest predicted position. If no valid association is found, *i.e.*, error of predicted position and cluster position is too large, the tracker re-clusters the raw points around the predicted position ${}^{G_i}\hat{\mathbf{p}}_m$. The reason is that the measured point reflectivity values are often affected by the object distance and laser incidence angle, so the extracted high-reflectivity points may not represent all points on the potential teammate drone. To obtain a more accurate cluster, we calculate a *predicted region* centered at ${}^{G_i}\hat{\mathbf{p}}_m$ and cluster an object within this region, which is then used to update the temporary tracker. Noted that the extracted objects whose size is much larger or smaller than actual drone size are rejected (invalid object). If no valid object can be clustered, the tracker will propagate to the next step. The tracker will be killed if it has propagated for too many steps without an update. Note that since the points in predicted region are far less than all input points, the time consumption of re-clustering would be significantly decreased.

2) Teammate Identification using Trajectory Matching: By rejecting invalid clusters and tracking real potential teammates, the trajectory of each temporary tracker is accumulated for subsequent identification. Since all drones in the proposed swarm system would exchange their estimated ego-state (in their own global frame) with others, the teammate identification and global extrinsic can be acquired by trajectory matching as follows:

$$\arg\min_{{}^{G_i}\mathbf{T}_{G_j}} \sum_{\kappa=1}^{\mathcal{K}} \frac{1}{2} \| {}^{G_i}\bar{\mathbf{p}}_{m,\kappa} - {}^{G_i}\mathbf{T}_{G_j}\,{}^{G_j}\breve{\mathbf{p}}_{b_j,\kappa} \| \tag{1}$$

where ${}^{G_i}\bar{\mathbf{p}}_{m,\kappa} \in {}^{G_i}\mathcal{T}_m$ represents the κ-th position in the tracked position trajectory ${}^{G_i}\mathcal{T}_m$ of the m-th object and ${}^{G_j}\breve{\mathbf{p}}_{b_j,\kappa} \in {}^{G_j}\mathcal{T}_j$ represents position received from drone j. Considering possible short-term communication disconnection, some data of ${}^{G_j}\breve{\mathbf{p}}_{b_j}$ might be lost. Thus, we only pick ${}^{G_i}\bar{\mathbf{p}}_{m,\kappa}$ that has close timestamp with ${}^{G_j}\breve{\mathbf{p}}_{b_j,\kappa}$ to participate in trajectory matching. Besides, to avoid large computing

[2]https://pcl.readthedocs.io/en/latest/cluster_extraction.html

Authorized licensed use limited to: ShanghaiTech University. Downloaded on October 05,2024 at 02:37:33 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: Initialization

Input: Raw LiDAR points in body frame ${}^{b_i}\mathcal{P}_{all}$, drone i's odometry ${}^{G_i}\mathbf{T}_{b_i}$, time interval of LiDAR input Δt, drone j's position trajectory ${}^{G_j}\mathcal{T}_j$, given threshold of trajectory matching residual thr.

Output: Global extrinsic transformation ${}^{G_i}\mathbf{T}_{G_j}$, tracked position of drone j in drone i's global frame ${}^{G_i}\mathbf{p}_{b_j}$

1 ${}^{b_i}\mathcal{P}_h$ = ReflectivityFiltering(${}^{b_i}\mathcal{P}_{all}$);
2 ${}^{b_i}\breve{\mathbf{p}}_m$ = EuclideanClustering(${}^{b_i}\mathcal{P}_h$);
3 **for** $m = 1 : M$ **do**
4 ${}^{G_i}\bar{\mathbf{p}}_m$ = TemporaryTracker(Δt, ${}^{G_i}\mathbf{T}_{b_i}$, ${}^{b_i}\mathcal{P}_{all}$, ${}^{b_i}\breve{\mathbf{p}}_m$);
5 ${}^{G_i}\mathcal{T}_m$.PushBack(${}^{G_i}\bar{\mathbf{p}}_m$);
6 **if** *TrajExcited*(${}^{G_i}\mathcal{T}_m$) **then**
7 **for** $= 1 : N$; $j \neq i$ **do**
8 res, $\mathbf{T}$ = TrajMatching(${}^{G_j}\mathcal{T}_j$, ${}^{G_i}\mathcal{T}_m$);
9 **if** $res < thr$ **then**
10 ${}^{G_i}\mathbf{T}_{G_j} = \mathbf{T}$;
11 ${}^{G_i}\mathbf{p}_{b_j}$ = ${}^{G_i}\mathbf{p}_m$;
12 break;
13 **end**
14 **end**
15 **end**
16 **end**

time due to too much data, we use a sliding window of the most recent $\mathcal{K}$ positions for matching.

Since no unique transformation can be determined from (1) if the involved trajectories are straight lines [28], the trajectories of those tracked objects are constantly assessed by **TrajExcited**(${}^{G_i}\mathcal{T}_m$) in Line 6 until sufficient information is collected. Let ${}^{G_i}\bar{\mathbf{p}}^c_m$ represent the centroid of ${}^{G_i}\mathcal{T}_m$, **TrajExcited**(${}^{G_i}\mathcal{T}_m$) assess the excitation (shape) of ${}^{G_i}\mathcal{T}_m$ by computing the singular values of matrix $\mathcal{H} \in \mathbb{R}^{3\times3}$:

$$\mathcal{H} \triangleq \sum_{\kappa=1}^{\mathcal{K}} ({}^{G_i}\bar{\mathbf{p}}_{m,\kappa} - {}^{G_i}\bar{\mathbf{p}}^c_m) \cdot ({}^{G_i}\bar{\mathbf{p}}_{m,\kappa} - {}^{G_i}\bar{\mathbf{p}}^c_m)^T \tag{2}$$

If the second largest singular value is larger than a given threshold, the trajectory is fully excited, which is qualified to do **TrajMatching**(${}^{G_j}\mathcal{T}_j$, ${}^{G_i}\mathcal{T}_m$) in Line 8. This function solves (1) which has well-studied closed-form solution [28]. The matching is performed with each received teammate drone's trajectory until the matching error is smaller than a given threshold, indicating that the object m is essentially the observation of teammate j, and the solution of (1) gives an initial estimation of the global extrinsic ${}^{G_i}\mathbf{T}_{G_j}$, which is then refined online using the ESIKF in Sec. III-D.3. After identification, a temporary tracker becomes a teammate tracker with the corresponding drone ID, which will be sequentially tracked as Sec. III-D.1. The initialization pipeline is illustrated in Fig. 3.

D. Decentralized LiDAR-inertial State Estimation

The fully decentralized state estimation of the proposed swarm system is a tightly-couple iterated Kalman filter inherited from FAST-LIO2 [4], but further incorporates the mutual observation constraints to improve ego-state estimation accuracy, and includes online refinement of inter-drone global extrinsic transformations.

1) Teammate Tracking: After the detection and identification of a teammate drone, a teammate tracker is obtained and the global extrinsic is initially calibrated by trajectory

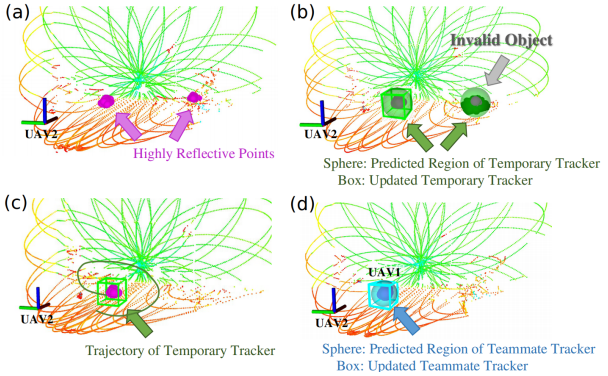

Fig. 3. Illustration of the initialization for newly detected objects, point-cloud is colored by reflectivity. Here the self drone is UAV2 and it needs to detect and identify other teammate UAVs in its FoV. Center of the box represents the updated position of the tracker. (a) Reflectivity filtering. (b) Outlier rejection by discarding objects with too large size. (c) Track real potential teammates and accumulate the trajectory. (d) After trajectory matching, the object is identified as UAV1, and the temporary tracker becomes a teammate tracker (see Sec. III-D.1).

matching. The teammate tracker is similar to a temporary tracker, but with two key differences. The first difference lies in the prediction. Both temporary and teammate tracker predict state using constant velocity model, but a teammate tracker predicts the state based on the velocity received from the corresponding teammate (and the most recent estimation of the corresponding global extrinsic transformation), instead of from ego-estimated velocity in a temporary tracker (see Sec. III-C.1). The second difference lies in state update when no teammate observation is available. In a temporary tracker, if there is no valid observation, *e.g.*, the teammate is outside FoV, the tracker will propagate for a few steps and then terminate as explained in Sec. III-C.1. In a teammate tracker, it will use the teammate odometry received from the network, after transforming to the self-drone's global frame using the most recent global extrinsic transformation obtained by Sec. III-D.3, to continue the state update.

2) State Prediction: Denote τ as the IMU measurement index, the discrete state transition model is shown below:

$$\mathbf{x}_{i,\tau+1} = \mathbf{x}_{i,\tau} \boxplus (\Delta t_\tau \mathbf{f}_i(\mathbf{x}_{i,\tau}, \mathbf{u}_{i,\tau}, \mathbf{w}_{i,\tau})) \tag{3}$$

Δt_τ is the time interval of two consecutive IMU measurements, $\mathbf{x}_{i,\tau}$ denotes ground-truth of state at IMU measurement's timestamp $t_{i\tau}$. The state vector $\mathbf{x}_i$, discrete state transition function $\mathbf{f}_i$, noise $\mathbf{w}_i$ and input $\mathbf{u}_i$ are defined as:

$$\begin{aligned}
\mathbf{x}_i &\triangleq [{}^{G_i}\mathbf{R}_{b_i}\ {}^{G_i}\mathbf{p}_{b_i}\ {}^{G_i}\mathbf{v}_{b_i}\ \mathbf{b}_{g_i}\ \mathbf{b}_{a_i}\ {}^{G_i}\mathbf{g} \\
&\quad \cdots\ {}^{G_i}\mathbf{R}_{G_j}\ {}^{G_i}\mathbf{p}_{G_j}\ \cdots] \in \mathcal{M} \\
\mathbf{f}_i &\triangleq [\omega_{m_i} - \mathbf{b}_{g_i} - \mathbf{n}_{g_i}\ {}^{G_i}\mathbf{v}_{b_i}\ {}^{G_i}\mathbf{R}_{b_i}(\mathbf{a}_{m_i} - \mathbf{b}_{a_i} - \mathbf{n}_{a_i}) + {}^{G_i}\mathbf{g} \\
&\quad \mathbf{n}_{\mathbf{b}_{gi}}\ \mathbf{n}_{\mathbf{b}_{ai}}\ \mathbf{0}_{3\times1}\ \cdots\ \mathbf{n}_R\ \mathbf{n}_p\ \cdots] \\
\mathbf{w}_i &\triangleq [\mathbf{n}_{g_i}\ \mathbf{n}_{a_i}\ \mathbf{n}_{\mathbf{b}_{gi}}\ \mathbf{n}_{\mathbf{b}_{ai}}\ \mathbf{n}_R\ \mathbf{n}_p], \quad \mathbf{u}_i \triangleq [\omega_{m_i}\ \mathbf{a}_{m_i}]
\end{aligned}$$

where $\omega_{m_i}, \mathbf{a}_{m_i}$ represent the IMU measurements of drone i, the meaning of each element in state vector $\mathbf{x}_i$ is introduced in Sec. III-A, the state manifold $\mathcal{M}$ are defined in (4) and its dimension is $18 + 6 \times (N-1)$.

$$\mathcal{M} \triangleq \underbrace{SO(3) \times \mathbb{R}^{15}}_{\dim=18} \times \underbrace{\cdots \times SO(3) \times \mathbb{R}^3 \times \cdots}_{\dim=6\times(N-1)} \tag{4}$$

Authorized licensed use limited to: ShanghaiTech University. Downloaded on October 05,2024 at 02:37:33 UTC from IEEE Xplore. Restrictions apply.

In (3), we used the notation $\boxplus/\boxminus$ defined in [29] to compactly represent the "plus" on the state manifold. Specifically, for the state manifold $SO(3)\times\mathbb{R}^n$ in (III-D.2), the $\boxplus$ operation and its inverse $\boxminus$ are defined as

$$\begin{bmatrix}\mathbf{R}\\ \mathbf{a}\end{bmatrix}\boxplus\begin{bmatrix}\mathbf{r}\\ \mathbf{b}\end{bmatrix}=\begin{bmatrix}\mathbf{R}\mathrm{Exp}(\mathbf{r})\\ \mathbf{a}+\mathbf{b}\end{bmatrix};\begin{bmatrix}\mathbf{R}_1\\ \mathbf{a}\end{bmatrix}\boxminus\begin{bmatrix}\mathbf{R}_2\\ \mathbf{b}\end{bmatrix}=\begin{bmatrix}\mathrm{Log}(\mathbf{R}_2^T\mathbf{R}_1)\\ \mathbf{a}-\mathbf{b}\end{bmatrix}$$

where $\mathbf{R},\mathbf{R}_1,\mathbf{R}_2\in SO(3),\mathbf{r}\in\mathbb{R}^3,\mathbf{a},\mathbf{b}\in\mathbb{R}^n$, $\mathrm{Exp}(\cdot):\mathbb{R}^3\mapsto SO(3)$ is the exponential map on $SO(3)$ [29] and $\mathrm{Log}(\cdot):SO(3)\mapsto\mathbb{R}^3$ is its inverse logarithmic map.

The state prediction step of the i-th drone's state under ESIKF framework is implemented once receiving a new IMU measurement as follows:

$$\widehat{\mathbf{x}}_{i,\tau+1}=\widehat{\mathbf{x}}_{i,\tau}\boxplus(\Delta t_\tau\mathbf{f}_i(\widehat{\mathbf{x}}_{i,\tau},\mathbf{u}_{i,\tau},\mathbf{0}));\widehat{\mathbf{x}}_0=\bar{\mathbf{x}}_{i,k-1}\quad(5)$$

3) Error State Iterative State Update: The update step is implemented iteratively at the end time of new LiDAR scan t_{ik}, fusing point-cloud measurements and mutual observation measurements (if any). Once receiving a new scan, motion compensation would be performed to obtain undistorted points, and point-to-plane distance will be computed to generate point-cloud residuals. Details of the motion compensation can be referred to [4]. Denote each motion undistorted point projected into global frame using propagated ego-state as ${}^{G_i}\widehat{\mathbf{p}}_n$, denote $\mathbf{u}_n$ as the normal vector of the corresponding plane, on which lying a point ${}^{G_i}\mathbf{q}_n$, the point residual is denoted as $\mathbf{z}_{p,n}$.

Apart from point-cloud residual, one main contribution of this paper is 3D LiDAR-based mutual observation measurements, which are used to construct new constraints to improve state estimation accuracy and to render the swarm system robust to degenerate scenes. For drone i, denote $\mathbf{z}_{ao,ij}$ as the active observation residual arising from ${}^{b_i}\breve{\mathbf{p}}_{b_j}$ (the active observation measurement w.r.t drone j, see Sec. III-D.1), and denote $\mathbf{z}_{po,ij}$ as the passive observation residual arising from ${}^{b_j}\breve{\mathbf{p}}_{b_i}$ (the passive observation measurement w.r.t drone j), then the residual block of drone i is composed as:

$$\begin{aligned}\mathbf{z}_i&=[\cdots,\mathbf{z}_{p,n}^T,\cdots,\mathbf{z}_{ao,ij}^T,\cdots,\mathbf{z}_{po,ij}^T,\cdots]^T\\ \mathbf{z}_{p,n}&=\mathbf{u}_n^T({}^{G_i}\widehat{\mathbf{p}}_n-{}^{G_i}\mathbf{q}_n)\\ \mathbf{z}_{ao,ij}&={}^{G_i}\mathbf{T}_{b_i}^{-1}{}^{G_i}\mathbf{T}_{G_j}{}^{G_j}\breve{\mathbf{p}}_{b_j}-{}^{b_i}\breve{\mathbf{p}}_{b_j}\\ \mathbf{z}_{po,ij}&={}^{G_j}\breve{\mathbf{T}}_{b_j}^{-1}{}^{G_i}\mathbf{T}_{G_j}^{-1}{}^{G_i}\mathbf{p}_{b_i}-{}^{b_j}\breve{\mathbf{p}}_{b_i}\end{aligned}\quad(6)$$

The state will be iteratively updated until convergence, the details can be referred to [4]. After convergence, the map will be incrementally updated using ikd-tree [30], an efficient data structure for map management. The relative state estimation is completed by projecting the ego-state information transmitted by teammate drones using the updated global extrinsic transformations.

IV. Experiments

The experiment platform contains three drones, and each one carries a Livox LiDAR (drone 1 and drone 3 carry Livox Mid360 and drone 2 carries Livox Avia with a smaller FoV), a Pixhawk flight controller (BMI055 6-axis IMU inside), an onboard computer Intel NUC with CPU i7-10710U. Each drone is attached with reflective tapes for easy detection (See Fig. 4). For each drone, the time offset and extrinsic between LiDAR and IMU are calibrated by [31], and the time among different drones are synced beforehand based on Network Time Protocol (NTP). All drones fly in manual mode using model predictive control [11].

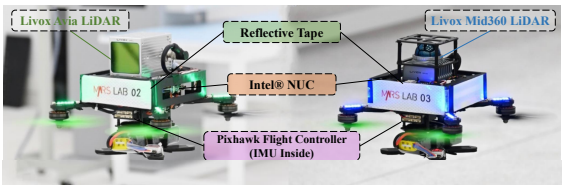

Fig. 4. The UAV platform of the proposed swarm system. Since the device setup of drone 1 is the same as drone 3, only the picture of drone 3 is shown.

A. Indoor Flights

Five flights in indoor scenes are performed to evaluate the localization accuracy. Motion capture system is introduced as the ground-truth. As far as we know, since there is no other 3D LiDAR-based state estimation methods for aerial swarm systems, we conduct ablation study by comparing localization accuracy of our method to FAST-LIO2 [4] (state-of-the-art LiDAR-inertial odometry for single drone system). From the position RMSE shown in Table I, it can be seen that for drone 2, the localization error of the proposed method fusing mutual observation measurements is obviously smaller than that of FAST-LIO2. The reason lies in the limited FoV of Livox Avia LiDAR, which is only $70.4^\circ\times 77.2^\circ$. The small FoV makes the LiDAR easier to drift due to lack of structural features, especially in indoor scenes. In this case, the mutual observation measurements can provide strong constraints to acquire accurate localization results. While for drone 1 and drone 3, thanks to the 360° horizontal FoV of Mid360 LiDAR, there are sufficient point-cloud constraints to keep FAST-LIO2 work well. Thus, fusing mutual observations does not improve the localization accuracy significantly. For all experiments the time interval of LiDAR input is $100\,\mathrm{ms}$, the average time-consumption per scan of the proposed method and FAST-LIO2 is $17.5\,\mathrm{ms}$ and $14.8\,\mathrm{ms}$. It proved the proposed Swarm-LIO can be implemented in real-time and has approximate computational efficiency to FAST-LIO2 (a fast odometry for single UAV system).

B. Degenerate Environment Flights

The first degenerate scene is data 05 shown in Table I, where drone 2 carrying a Livox Avia LiDAR faces a smooth wall which imposes insufficient constraints to determine the full state. As a result, the single-agent odometry FAST-LIO2 completely degenerates with large RMSE. While for the proposed swarm system, drone 2 is observed by its teammate drone 1 and drone 3 (the passive observation measurement). Fusing passive observation measurements transmitted from drone 1 and drone 3 in (6) will still lead to stable and accurate state estimation. The point-cloud map, mutual observation measurements are illustrated in Fig. 5.

The second degenerate scene is a smooth cuboid corridor. Similar to the smooth wall, observations of a single LiDAR

Authorized licensed use limited to: ShanghaiTech University. Downloaded on October 05,2024 at 02:37:33 UTC from IEEE Xplore. Restrictions apply.

TABLE I

LOCALIZATION ACCURACY EVALUATION

Data	Drone 1		Drone 2		Drone 3	
	Swarm-LIO	FAST-LIO2	Swarm-LIO	FAST-LIO2	Swarm-LIO	FAST-LIO2
01	0.0407	**0.0399**	**0.0464**	0.0673	0.0359	**0.0347**
02	0.0296	**0.0289**	**0.0428**	0.0692	**0.0305**	0.0307
03	**0.0303**	0.0318	**0.0442**	0.0836	**0.0300**	0.0313
04	0.0355	**0.0341**	**0.0412**	0.0827	**0.0298**	0.0310
05	**0.0242**	0.0255	**0.0536**	3.0844	0.0327	**0.0319**

The position error (RMSE) of FAST-LIO2 and the proposed Swarm-LIO (unit: meter).

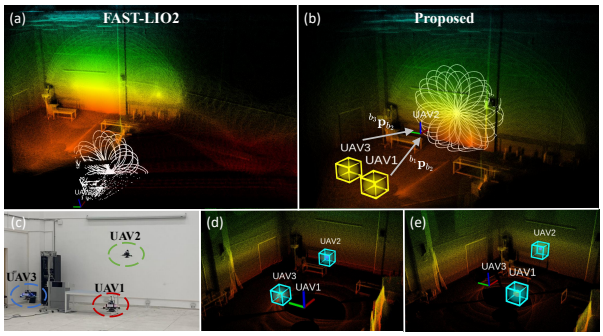

Fig. 5. Degenerate smooth wall experiment. The white points in (a) and (b) refer to current input points. (a) View of drone 2 with point-cloud map constructed by FAST-LIO2, severe drift occurs and the map is messy. (b) View of drone 2 with point-cloud map constructed by the proposed method fusing passive observation measurements ${}^{b_1}\mathbf{p}_{b_2}$ and ${}^{b_3}\mathbf{p}_{b_2}$. Yellow boxes means that teammate drones are out of FoV, teammate trackers are updated fusing ego-state estimation transmitted by teammates. (c) Picture of the three drones. (d)(e) View of drone 1 and drone 3 respectively and their observations of teammate drones (the blue boxes).

do not provide sufficient constraints for pose determination, so FAST-LIO2 suffers from drifts. While for a swarm system, when the first drone flies into the corridor, other drones still hover at the entrance (where enough structural features exist for their state estimation) and provide passive observation measurements for the first drone. After the first drone flies through the corridor, it hovers at the end, offering passive observation measurements for rest of the drones, so they could also fly through the corridor. Thus, the whole swarm can fly through a smooth corridor without any degeneration. From Fig. 6, it can be seen the proposed method is robust to this type of degenerate scene, the map can recover more structural details than FAST-LIO2. More visual illustrations can be found in our video[3].

C. Decentralized Exploration

To validate that the proposed swarm system is robust to observation lost and can calibrate the global extrinsic transformations, a large-scale decentralized exploration experiment is conducted. Three drones are randomly placed in front of a mansion and then take off. After the take-off, they perform some random flights during which each drone detects, tracks, and identifies each other all automatically and without any prior extrinsic information. After that the drones fly in different directions to explore different areas. After flights finish, the point-cloud map of each drone can be

[3]https://youtu.be/MxeoKVXrmEs

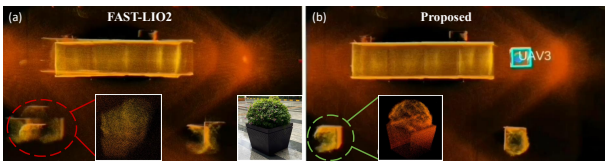

Fig. 6. Point-cloud map of smooth cuboid corridor scene constructed by FAST-LIO2 (a) and the proposed method (b).

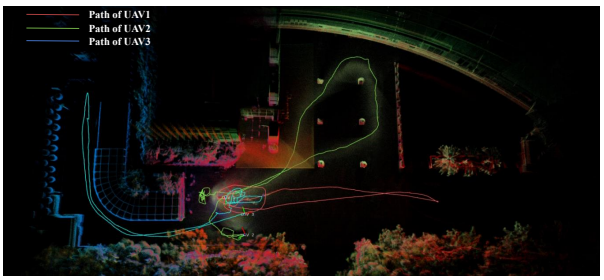

Fig. 7. Merged point-cloud map. The red, green, blue points are scanned by drone 1, drone2, drone 3 respectively.

merged together offline using the estimated global extrinsic. By transforming point-cloud map of drone 2 and drone 3 to drone 1's global frame, the merged map is shown in Fig. 7. The consistent merged map shows the capability of accurate global extrinsic estimation without any initial value. Another point should be noted is, since the drones fly in different regions, there is no mutual observation for a long time, *e.g.*, for drone 2 the average distance without mutual observation is $88.58\,\mathrm{m}$. But still, when the teammate drone returns to drone 2's FoV, the error between the detected teammate position and the predicted position by the teammate tracker (based on the received teammate ego-state, see Sec. III-D.1) is only $0.17\,\mathrm{m}$, which indicates the high accuracy of the estimated global extrinsic transformation. More visual illustrations are shown in our video.

V. CONCLUSION AND FUTURE WORK

This paper proposed a fully decentralized and accurate swarm LiDAR-inertial odometry. A novel 3D LiDAR-based drone detection, identification and tracking method was proposed to perform fully autonomous initialization and robust relative state estimation. Mutual observation measurements are tightly-coupled with IMU and point-cloud measurements under ESIKF framework, fulfilling real-time and accurate ego-state estimation, which can keep robust even in degenerate environments for camera or LiDAR. In the future, drone detection can be upgraded to an auxiliary-free method (*e.g.*, no reflective tape), and the initialization can be more effective. Besides, the scale of the swarm can be larger, more drones will be produced to complete various tasks.

VI. ACKNOWLEDGEMENT

This work is supported in part by the University Grants Committee of Hong Kong General Research Fund under Project 17206421 and in part by DJI under the grant 200009538. The authors gratefully acknowledge Livox Technology for the equipment support and thank Dr. Ximin Lyu of Sun Yat-sen University for the experimental site support during the whole work.

Authorized licensed use limited to: ShanghaiTech University. Downloaded on October 05,2024 at 02:37:33 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] Yuman Gao, Yingjian Wang, Xingguang Zhong, Tiankai Yang, Mingyang Wang, Zhixiong Xu, Yongchao Wang, Yi Lin, Chao Xu, and Fei Gao. Meeting-merging-mission: A multi-robot coordinate framework for large-scale communication-limited exploration. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 13700–13707. IEEE, 2022.

[2] Boyu Zhou, Yichen Zhang, Xinyi Chen, and Shaojie Shen. Fuel: Fast uav exploration using incremental frontier structure and hierarchical planning. *IEEE Robotics and Automation Letters*, 6(2):779–786, 2021.

[3] Pengfei Zhu, Jiayu Zheng, Dawei Du, Longyin Wen, Yiming Sun, and Qinghua Hu. Multi-drone-based single object tracking with agent sharing network. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(10):4058–4070, 2020.

[4] Wei Xu, Yixi Cai, Dongjiao He, Jiarong Lin, and Fu Zhang. Fast-lio2: Fast direct lidar-inertial odometry. *IEEE Transactions on Robotics*, 2022.

[5] Tong Qin, Peiliang Li, and Shaojie Shen. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4):1004–1020, 2018.

[6] Jiarong Lin and Fu Zhang. R3live: A robust, real-time, rgb-colored, lidar-inertial-visual tightly-coupled state estimation and mapping package. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 10672–10678. IEEE, 2022.

[7] Jiarong Lin and Fu Zhang. R^3live++: A robust, real-time, radiance reconstruction package with a tightly-coupled lidar-inertial-visual state estimator. *arXiv preprint arXiv:2209.03666*, 2022.

[8] Yunfan Ren, Fangcheng Zhu, Wenyi Liu, Zhepei Wang, Yi Lin, Fei Gao, and Fu Zhang. Bubble planner: Planning high-speed smooth quadrotor trajectories using receding corridors. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6332–6339. IEEE, 2022.

[9] Fanze Kong, Wei Xu, Yixi Cai, and Fu Zhang. Avoiding dynamic small obstacles with onboard sensing and computation on aerial robots. *IEEE Robotics and Automation Letters*, 6(4):7869–7876, 2021.

[10] Yunfan Ren, Siqi Liang, Fangcheng Zhu, Guozheng Lu, and Fu Zhang. Online whole-body motion planning for quadrotor using multi-resolution search. *arXiv preprint arXiv:2209.06761*, 2022.

[11] Guozheng Lu, Wei Xu, and Fu Zhang. On-manifold model predictive control for trajectory tracking on robotic systems. *IEEE Transactions on Industrial Electronics*, 2022.

[12] Hao Xu, Luqi Wang, Yichen Zhang, Kejie Qiu, and Shaojie Shen. Decentralized visual-inertial-uwb fusion for relative state estimation of aerial swarm. In *2020 IEEE international conference on robotics and automation (ICRA)*, pages 8776–8782. IEEE, 2020.

[13] Xin Zhou, Xiangyong Wen, Zhepei Wang, Yuman Gao, Haojia Li, Qianhao Wang, Tiankai Yang, Haojian Lu, Yanjun Cao, Chao Xu, et al. Swarm of micro flying robots in the wild. *Science Robotics*, 7(66):eabm5954, 2022.

[14] Xin Zhou, Jiangchao Zhu, Hongyu Zhou, Chao Xu, and Fei Gao. Ego-swarm: A fully autonomous and decentralized quadrotor swarm system in cluttered environments. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4101–4107. IEEE, 2021.

[15] Wolfgang Hönig, James A Preiss, TK Satish Kumar, Gaurav S Sukhatme, and Nora Ayanian. Trajectory planning for quadrotor swarms. *IEEE Transactions on Robotics*, 34(4):856–869, 2018.

[16] Aldo Jaimes, Srinath Kota, and Jose Gomez. An approach to surveillance an area using swarm of fixed wing and quad-rotor unmanned aerial vehicles uav (s). In *2008 IEEE International Conference on System of Systems Engineering*, pages 1–6. IEEE, 2008.

[17] SungTae Moon, YeonJu Choi, DoYoon Kim, Myeonghun Seung, and HyeonCheol Gong. Outdoor swarm flight system based on rtk-gps. *Journal of KIISE*, 43(12):1315–1324, 2016.

[18] Pierre-Yves Lajoie, Benjamin Ramtoula, Yun Chang, Luca Carlone, and Giovanni Beltrame. Door-slam: Distributed, online, and outlier resilient slam for robotic teams. *IEEE Robotics and Automation Letters*, 5(2):1656–1663, 2020.

[19] Renaud Dubé, Abel Gawel, Hannes Sommer, Juan Nieto, Roland Siegwart, and Cesar Cadena. An online multi-robot slam system for 3d lidars. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1004–1011. IEEE, 2017.

[20] Hao Xu, Yichen Zhang, Boyu Zhou, Luqi Wang, Xinjie Yao, Guotao Meng, and Shaojie Shen. Omni-swarm: A decentralized omnidirectional visual–inertial–uwb state estimation system for aerial swarms. *IEEE Transactions on Robotics*, 2022.

[21] Haoyu Zhou, Zheng Yao, Zhuo Zhang, Penghao Liu, and Mingquan Lu. An online multi-robot slam system based on lidar/uwb fusion. *IEEE Sensors Journal*, 22(3):2530–2542, 2021.

[22] Dongjiao He, Wei Xu, and Fu Zhang. Kalman filters on differentiable manifolds. *arXiv preprint arXiv:2102.03804*, 2021.

[23] Aaron Weinstein, Adam Cho, Giuseppe Loianno, and Vijay Kumar. Visual inertial odometry swarm: An autonomous swarm of vision-based quadrotors. *IEEE Robotics and Automation Letters*, 3(3):1801–1807, 2018.

[24] Ty Nguyen, Kartik Mohta, Camillo J Taylor, and Vijay Kumar. Vision-based multi-mav localization with anonymous relative measurements using coupled probabilistic data association filter. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3349–3355. IEEE, 2020.

[25] Christopher E Denniston, Yun Chang, Andrzej Reinke, Kamak Ebadi, Gaurav S Sukhatme, Luca Carlone, Benjamin Morrell, and Ali-akbar Agha-mohammadi. Loop closure prioritization for efficient and scalable multi-robot slam. *arXiv preprint arXiv:2205.12402*, 2022.

[26] Yewei Huang, Tixiao Shan, Fanfei Chen, and Brendan Englot. Disco-slam: Distributed scan context-enabled multi-robot lidar slam with two-stage global-local graph optimization. *IEEE Robotics and Automation Letters*, 7(2):1150–1157, 2021.

[27] Alicja Wasik, Rodrigo Ventura, José N Pereira, Pedro U Lima, and Alcherio Martinoli. Lidar-based relative position estimation and tracking for multi-robot systems. In *Robot 2015: Second Iberian Robotics Conference*, pages 3–16. Springer, 2016.

[28] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(5):698–700, 1987.

[29] C. Hertzberg, R. Wagner, U. Frese, and L. Schröder. Hertzberg, christoph and wagner, rené and frese, udo and schröder, lutz. *Information Fusion*, 14(1):57–77, 2013.

[30] Yixi Cai, Wei Xu, and Fu Zhang. ikd-tree: An incremental kd tree for robotic applications. *arXiv preprint arXiv:2102.10808*, 2021.

[31] Fangcheng Zhu, Yunfan Ren, and Fu Zhang. Robust real-time lidar-inertial initialization. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3948–3955. IEEE, 2022.

Authorized licensed use limited to: ShanghaiTech University. Downloaded on October 05,2024 at 02:37:33 UTC from IEEE Xplore. Restrictions apply.