HCCNet: Hybrid Coupled Cooperative Network for Robust Indoor Localization

LI ZHANG, College of Mathematics, Hefei University of Technology, Hefei, China
XU ZHOU, Hefei University of Technology, Hefei, China
DANYANG LI, Tsinghua University, Beijing, China
ZHENG YANG, School of Software, Tsinghua University, Beijing, China

Accurate localization of unmanned aerial vehicle (UAV) is critical for navigation in GPS-denied regions, which remains a highly challenging topic in recent research. This article describes a novel approach to multi-sensor hybrid coupled cooperative localization network (HCCNet) system that combines multiple types of sensors including camera, ultra-wideband (UWB), and inertial measurement unit (IMU) to address this challenge. The camera and IMU can automatically determine the position of UAV based on the perception of surrounding environments and their own measurement data. The UWB node and the UWB wireless sensor network (WSN) in indoor environments jointly determine the global position of UAV, and the proposed dynamic random sample consensus (D-RANSAC) algorithm can optimize UWB localization accuracy. To fully exploit UWB localization results, we provide an HCCNet system which combines the local pose estimator of visual inertial odometry (VIO) system with global constraints from UWB localization results. Experimental results show that the proposed D-RANSAC algorithm can achieve better accuracy than other UWB-based algorithms. The effectiveness of the proposed HCCNet method is verified by a mobile robot in real world and some simulation experiments in indoor environments.

CCS Concepts: • **Networks → Location based services;**

Additional Key Words and Phrases: Indoor localization, VIO, UWB, dynamic RANSAC, hybrid coupling, multi-sensor network

ACM Reference Format:
Li Zhang, Xu Zhou, Danyang Li, and Zheng Yang. 2024. HCCNet: Hybrid Coupled Cooperative Network for Robust Indoor Localization. *ACM Trans. Sensor Netw.* 20, 4, Article 100 (July 2024), 22 pages. https://doi.org/10.1145/3665645

1 INTRODUCTION

Unmanned aerial vehicles (UAVs) have been widely applied in various fields, including civilian applications [22], target tracking [47], search and rescue [1], and industrial inspections [16], demonstrating immense potential. In indoor applications, real-time localization of UAVs is a critical

This work was supported in part by the National Science Foundation of China under Grant 61972131 and in part by the National Key Research and Development Program of China under Grant 2018YFB2100301.
Authors' Contact Information: Li Zhang (Corresponding author), College of Mathematics, Hefei University of Technology, Hefei, Anhui, China; e-mail: lizhang@hfut.edu.cn; Xu Zhou, Hefei University of Technology, Hefei, Anhui, China; e-mail: zhouxu4695@gmail.com; Danyang Li, Tsinghua University, Beijing, Beijing, China; e-mail: lidanyang1919@gmail.com; Zheng Yang, School of Software, Tsinghua University, Beijing, China; e-mail: hmilyyz@gmail.com.

ACM 1550-4859/2024/07-ART100
https://doi.org/10.1145/3665645

component of **guidance, navigation, and control (GNC)** systems. Almost all UAVs are equipped with **Global Navigation Satellite System (GNSS)** receivers for outdoor localization. However, GNSS is unreliable or even unavailable in indoor environments due to signal attenuation caused by buildings, walls and other structures. To overcome the complexity of indoor environments and the limitations of GNSS unavailability, researchers have proposed various innovative localization methods and technologies, such as WiFi [35], RFID [46], Bluetooth [48], LiDAR [38], and inertial navigation [42].

Visual inertial odometry (VIO), combining visual and IMU data, is becoming a key method for indoor UAV localization. It uses cameras to capture and analyze environmental images, identifying feature keypoints [24] and tracking optical flow [33]. Simultaneously, **inertial measurement unit (IMU)** measures the UAV's acceleration and angular velocity. This integration enhances the robustness and accuracy of UAV localization and scene understanding. Although VIO technology is one of the primary methods for indoor UAV localization, it still faces two main challenges: accumulated drift and the limitation of global coordinate localization. Accumulated drift is a common issue in VIO methods, where scale information gradually deviates from the real-world scale due to camera and IMU measurement errors, inaccurate feature tracking, and other factors. This accumulated drift results in a deviation between localization results and real environments, which becomes more prominent in long-duration operations or large-scale movements. In addition, VIO methods focus on estimating a camera's relative coordinate position changes but do not directly provide absolute localization information about the global coordinate. These two problems limit the practical application of VIO methods.

Integrating **ultra-wideband (UWB)** sensors offers a promising remedy for tackling accumulated drift within VIO while establishing comprehensive global coordinate localization [15]. UWB technology's wideband signal transmission and precise ranging capabilities enable accurate distance measurement and position estimation. The accumulated drift issue in VIO can be rectified by installing UWB sensors on the UAV to measure the distance between the UAV and the ground base station. In addition, UWB sensors provide global coordinate position data by deploying multiple base stations in indoor environments. However, UWB sensors are affected by indoor environments which will lead the decrease in localization accuracy.

By reason of the foregoing, we propose an HCCNet system which employs a novel hybrid coupled method to fuse the results of UWB and VIO. We first introduce a D-RANSAC algorithm to improve the quality of UWB localization results in indoor environments. Subsequently, our methodology extends to the sophisticated integration of tightly and loosely coupled mechanisms within the HCCNet framework, aiming to optimize the fusion of multi-sensor localization data. Through extensive experiments, we demonstrate that D-RANSAC can effectively suppress UWB localization outliers and obtain high-quality localization results with a small number of anchors. The proposed HCCNet system significantly improves accuracy and robustness compared to the state-of-the-art VIO system by fully exploiting UWB localization results. Our tests show that our system can locate positions within about 9cm accuracy in various real settings.

The main contributions of this article are as follows:

(1) A novel outlier filtering technique, namely D-RANSAC algorithm, is proposed in this article. This technique combines the sliding window method and the RANSAC algorithm to identify and remove outliers effectively.
(2) The HCCNet system is introduced in this article. It integrates nonlinear optimization and particle filtering approaches to fuse the localization results of UWB and VIO. This phased approach mitigates the issue of accumulated drift over time and provides reliable global localization results.

(3) We conduct extensive simulations and real-world experiments to evaluate the performance of our system. The experimental results show that our proposed HCCNet system outperforms other state-of-the-art frameworks for localization accuracy.

2 RELATED WORK

2.1 Multi-sensor Fusion Localization Methods

Multi-sensor fusion algorithms in the field of localization can be broadly categorized into tightly coupled [13, 25, 27] and loosely coupled approaches [2, 23]. The loosely coupled framework fuses the pose estimation results of different sensors separately, which cannot fully use their respective advantages. However, in a tightly coupled framework, the information between different sensors and subsystems affects each other, and their advantages can be better used.

In loosely coupled approaches, the VIO module operates as an independent entity. These methods do not continuously integrate new measurements to improve real-time accuracy. Instead, they address accumulated drift intermittently, usually during infrequent location requests, leading to corrections by jumps. Qin et al. [32] and Mascaro et al. [21] employed pose graph optimization techniques to enhance the alignment between GPS and VIO coordinates. Gao et al. [7] introduced a method for low-drift VIO in indoor settings. They combined UWB localization with a cost function and nonlinear optimization to enhance location measurements. In a similar vein, Liu et al. [17] presented a technique for mobile robot localization and map construction. They fused UWB, odometry, and 2D lidar data and employed graph optimization to mitigate the effects of cumulative errors. Xu et al. [44] achieved instant indoor navigation with a high success rate by using the trajectory experience of previous facilitators to guide future users in a peer-to-peer mode. Lin et al. [15] improved the accuracy of VIO by integrating UWB localization using a loosely coupled approach for the localization problem of UAVs in GPS-denied environments. Gong et al. [8] used a two-stage loosely coupled adaptive fusion method to fuse GNSS and VIO for consistent and accurate global positioning in a GNSS intermittent degradation environment. Zhao et al. [49] proposed a loosely coupled localization system that integrates multi-base station UWB technology and point-line VIO. This approach improves localization accuracy and robustness by leveraging multi-sensor fusion capabilities. In a similar vein, Nguyen et al. [26] devised a scale correction technique for monocular **visual odometry (VO)** systems, using UWB distances for real-time scale estimation. This method does not rely on anchor location information and is compatible with different sensor types. However, the drawback of the loosely coupled method is that the mutual information between sensors cannot be fully utilized, leading to the loss of accuracy and robustness. In complex scenarios, loosely coupled methods may not provide sufficiently accurate estimation results.

In tightly coupled approaches, global localization poses or extra distances ranging are integrated as additional measurements. These approaches can enhance real-time accuracy within an **Extended Kalman Filter (EKF)** or an optimization-based framework. Nguyen et al. [28] put forward a tightly coupled approach that combines image, IMU, and UWB inputs. Magnago et al. [20] used an unscented Kalman filter to fuse relative control input and global information from UWB localization to mitigate cumulative errors in ground mobile robots. The **Monte Carlo localization (MCL)** technique was utilized by Perez-Grau et al. [31]. They integrated RGB-D point cloud data, UWB localization, and IMU data within this approach. Wang et al. [40] introduced a method that simplified loop detection using multiple pre-configured UWB anchors, reducing computation complexity and error accumulation. Cao and Beltrame [4] used UWB information to mitigate errors and enhanced mobility by incorporating a single unassigned anchor for localization while retaining the loop detection module. Yang et al. [45] creatively proposed a resilient multi-sensor fusion method to overcome sensor failures and complex scenes to achieve robust and accurate indoor

localization. Additionally, Zheng et al. [50] developed a UWB-VIO fusion system aimed at achieving precise and reliable relative localization for circular robotic teams. Their focus centered on refining the initialization approach and improving processing in **non-line-of-sight (NLOS)** conditions. Jung et al. [9] introduced U-VIO, an intricately combined UWB VIO algorithm that utilizes UWB measurements throughout both front-end and back-end graph optimization stages. Meanwhile, Choi et al. [5] brought forth LUVI, a streamlined UWB-VIO relative positioning method tailored for AR-IoT applications. They prioritize the efficient management of virtual anchors and computational speed. In another vein, Liu et al. [19] proposed an integrated algorithm for autonomous vehicle localization, merging GNSS, UWB, and VIO to boost accuracy and reliability, especially in intricate surroundings. This approach capitalizes on UWB technology for indoor absolute positioning and VIO for improved navigation in GNSS-obstructed environments. Lastly, Wang et al. [41] presented a localization scheme designed for range-assisted VIO systems. By incorporating UWB measurements, they effectively counter VIO drift. Their method features filter-based UWB anchor localization and observability-constrained optimization. This approach seamlessly blends visual, inertial, and UWB data, even when fewer than four UWB anchors are available. Although tightly coupled methods have higher accuracy and robustness, their complexity and sensor requirements may limit their scope in some applications.

2.2 UWB Localization Methods

Over the past decade, UWB localization has relied on **Time of Arrival (TOA)** or **Time Difference of Arrival (TDOA)** computation to determine the position of a tag relative to known reference locations called anchors. The high-speed transmission of UWB signals, strong anti-interference capabilities, and low-power consumption make them well suited for indoor localization systems.

Recently, several UWB-based approaches have emerged, particularly targeting environments where GNSS signals are unavailable. Bottigliero et al. [3] introduced an indoor **real-time locating system (RTLS)**. The system utilized measurements obtained from non-synchronized UWB pulse sequences. Liu et al. [18] introduced a cooperative localization system that combines UWB and Wi-Fi measurements with inertial data. This approach combined long-range Wi-Fi and short-range UWB localization for multiple users, followed by IMU-based dead reckoning refinement. Li et al. [14] proposed the ViViPlus indoor positioning system based on WiFi fingerprint, which successfully solves the spatial ambiguity and temporal instability problems by introducing a novel **RSS spatial gradient (RSG)** matrix form, and significantly improves the positioning accuracy without additional information or user constraints. Queralta et al. [34] suggested using only **Time of Flight (ToF)** distance estimates for localization, analyzing anchor self-correction for mobile robots to improve system mobility. Lin et al. [36] proposed an error minimization protocol for node localization in WSN. By combining the information of **received signal strength (RSS)** and TOA, a reliable estimation of the position of the node is achieved when the position of the anchor node is known. Another study by Nguyen et al. [29] investigated collaborative localization among drones and other platforms. Recent studies have also examined the impact of obstacles on UWB localization accuracy [43] and proposed composite filtering methods for handling dynamic uncertainty in quadrotor UAVs [11]. However, UWB localization results are still affected by the indoor environment.

In this article, we use a novel HCCNet system to combine VIO and UWB localization results. Because of UWB environmental interference, a D-RANSAC algorithm is proposed to filter localization outliers. Then, we combine tightly coupled and loosely coupled methods to solve the accumulated drift and obtain high-precision global localization results.

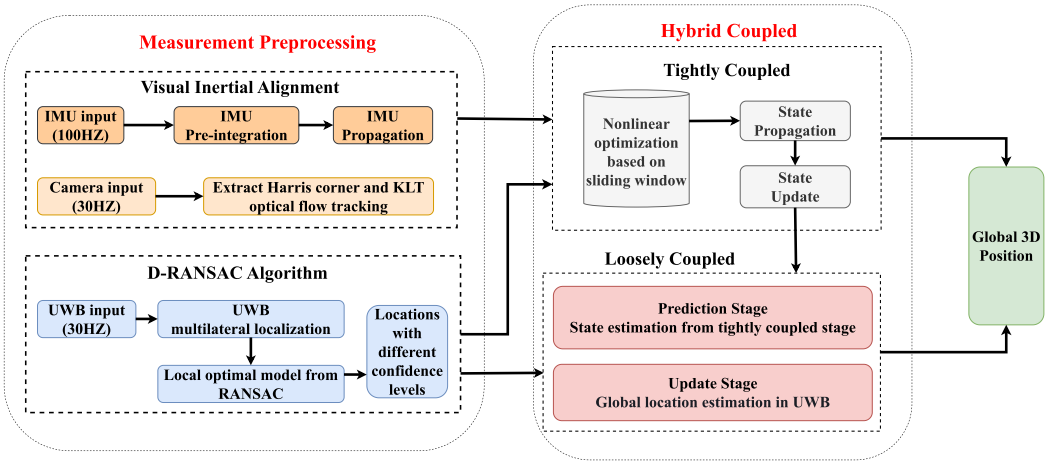


Fig. 1. Flowchart of HCCNet system.

3 SYSTEM OVERVIEW

The proposed HCCNet system is structured using a framework shown in Figure 1. It comprises two main components: a measurement preprocessing module and a hybrid coupled cooperative module.

In the measurement preprocessing module, the preprocessing process is performed on camera, IMU and UWB sensors. More specifically, the feature information in the image is extracted, and IMU data between two consecutive frames is pre-integrated. The UWB localization results are affected by outliers generated by indoor environments, which are filtered by the proposed D-RANSAC algorithm. The camera, IMU, and UWB measurements are then forwarded to a tightly coupled module for pose estimation.

In the hybrid coupled module, we first adopt a tightly coupled method to obtain preliminary pose. Then, a loosely coupled method based on particle filtering is used to fuse the tightly coupled pose and UWB global localization results. In this stage, the hybrid coupled module will achieve accurate pose estimation.

Overall, the proposed HCCNet system provides an efficient and accurate approach to estimate system position in real-time. The measurement preprocessing, tightly coupled nonlinear optimization, and loosely coupled particle filtering approaches provide a comprehensive solution to ensure reliable position estimation.

4 D-RANSAC ALGORITHM

RANSAC emerges in machine vision literature as a quick algorithm for outlier removal and camera pose estimation [6]. It excels in datasets filled with many outliers. Yet, it requires at least 50% of the data to be inliers for best results. This method aims to estimate model parameters using a small observed data subset. However, RANSAC might find it challenging to pick such a subset accurately. This subset needs to reflect complex motion patterns found in UAV motion trajectories' prevalent nonlinearities. In cases of nonlinear motion, outliers tend to scatter widely and unevenly. This scattering poses big challenges for RANSAC in precisely identifying and removing outliers. UWB localization outcomes intensify RANSAC's challenges due to signal propagation's complex dynamics. This complexity results in more noticeable outliers, increasing RANSAC's difficulty in this area.

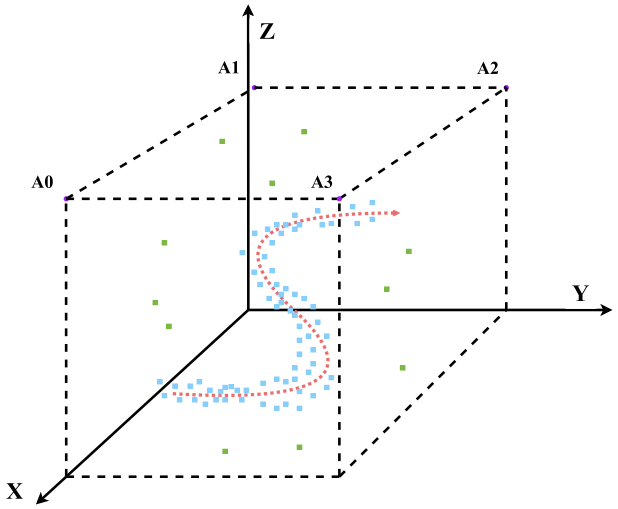


Fig. 2. Outliers removed by D-RANSAC algorithm.

In indoor environments, the influence of noise on UWB-based localization is inevitable, leading to potential inaccuracies in position estimation. This article assumes that UAV motion can be approximated as a straight line for a short period, which is common in various UAV applications. To remove outliers from UAV trajectories efficiently, we employ a brand new D-RANSAC algorithm which combines the advantages of the classic RANSAC algorithm and the sliding window method.

The measurement frequency of UWB sensor is used to select M consecutive locations to form short-term straight trajectories. The D-RANSAC algorithm dedicates a corresponding line model from these M locations. First, a line model is fitted to two selected sample locations, randomly. Next, the error between the all data samples and the model is calculated. Samples with a fitting error less than a preset threshold are considered interior points. Subsequently, two additional random sample locations are selected, and the process is repeated to obtain different models. The quality of each model is evaluated based on the number of interior points in the model. Finally, the optimal local model and its corresponding interior points are obtained through this process. Algorithm 1 shows the specific process of D-RANSAC.

Figure 2 illustrates a flight trajectory of a UAV. This trajectory comprises a sequence of position points, each serving a distinct purpose. The purple points represent UWB anchors, and the red points correspond to the real trajectory of the UAV. The blue and green points represent the position points calculated by UWB anchors, which provide auxiliary data for trajectory estimation. The D-RANSAC algorithm is used to identify green points and effectively remove orange outliers. After removing the orange points, the trajectory formed by the green points is closer to the real trajectory of the UAV.

The proposed methodology involves evaluating each location M times and assigning a confidence value based on the number of times it has been evaluated as an interior point. The confidence value is calculated by

$$\begin{aligned} \mathbf{w}_k &\propto \frac{\mathbf{p}(\mathbf{Z}_k|\mathbf{I}_k)\mathbf{p}(\mathbf{I}_k)}{\mathbf{p}(\mathbf{Z}_k)} \\ &= \frac{\mathbf{p}(\mathbf{I}_k)\prod_{i\neq k}\mathbf{p}(\mathbf{Z}_i|\mathbf{I}_i)}{\mathbf{p}(\mathbf{Z}_k|\mathbf{I}_k)\mathbf{p}(\mathbf{I}_k)+\mathbf{p}(\mathbf{Z}_k|\mathbf{O}_k)\mathbf{p}(\mathbf{O}_k)}, \end{aligned} \tag{1}$$

ALGORITHM 1: D-RANSAC.

Input: UWB sensor measurements, sliding window size M, distance threshold $dist$
Output: Local optimal model M_{opt}

```
Initialize local optimal model M_opt;
for sliding window do
    Initialize inlier set S_in and outlier set S_out;
    for i = 1 to N iterations do
        Randomly select 2 points to form a model;
        Compute the distance;
        Determine rationality of each point;
        if distance ≤ dist then
            Add point to inlier set S_in;
        else
            Add point to outlier set S_out;
        end
    end
    Evaluate model quality;
    Select local optimal model M_opt;
end
return M_opt;
```

where $\mathbf{I}_k$ and $\mathbf{O}_k$ denote the probability that point k is an inlier or outlier, respectively; $\mathbf{Z}_k$ denotes the distance from point k to the optical model. The values of i and k range from $0, 1, 2, \ldots, M$.

This approach ensures that the confidence of each location is proportional to the number of its interior evaluations, providing a measure of certainty. Furthermore, this confidence value is used in subsequent stages of the procedure, making the system resilient to outliers.

By incorporating confidence values into the approach, we can reduce the effect of outliers and improve the accuracy of the results. This strategy is particularly beneficial in scenarios where the data is susceptible to noise or other sources of variability. This approach can produce consistent and reliable results robust to outliers.

5 HYBRID COUPLED COOPERATIVE LOCALIZATION NETWORK

5.1 Tightly Coupled Stage

The hybrid coupled model, illustrated in Figure 3, employs distinct strategies at different stages. During the tightly coupled stage, nonlinear optimization comes into play, whereas the loosely coupled stage utilizes the particle filter method.

VIO is a robust tool for achieving real-time precision in both indoor and outdoor settings. However, a notable drawback emerges the accumulation of errors over time, resulting in significant localization drift issues during practical use. To counter this challenge, we propose a solution: introduce constraints to the VIO system via the short-term variations in UWB localization outcomes.

In the back-end optimization of our system, a tightly coupled nonlinear optimization method is mainly used. The corresponding residual terms are constructed to solve the maximum likelihood estimation problem based on the measurement data of different sensors. The best solution of optimization process can be seen as a precise state estimation. The states for estimation are defined as:

$$\mathbf{X} = [\mathbf{X}_U, \mathbf{X}_{VI}], \tag{2}$$

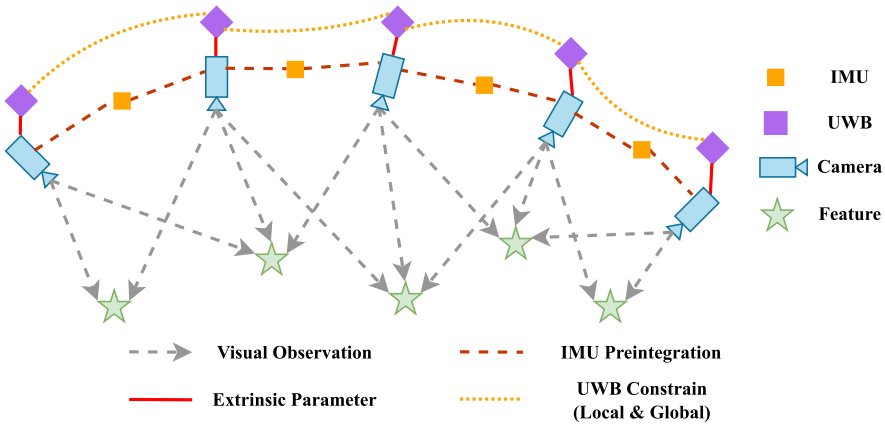


Fig. 3. Diagram of multi-sensor hybrid coupled fusion method.

$$\mathbf{X}_U = [\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_n], \tag{3}$$

$$\mathbf{X}_{VI} = [\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n, \lambda_0, \lambda_1, \ldots, \lambda_m], \tag{4}$$

$$\mathbf{x}_k = [\mathbf{p}^c_{bk}, \mathbf{v}^c_{bk}, \mathbf{q}^c_{bk}, \mathbf{b}_a, \mathbf{b}_g], k = 0, 1, 2, \ldots, n, \tag{5}$$

$$\mathbf{u}_h = [\mathbf{p}^G_{xh}, \mathbf{p}^G_{yh}, \mathbf{p}^G_{zh}], h = 0, 1, 2, \ldots, n. \tag{6}$$

The state vector is denoted as $\mathbf{X}$ with including the states of IMU, UWB, and vision. The kth IMU state, denoted as $\mathbf{x}_k$, which includes the position $\mathbf{p}^c_{bk}$ of the IMU center relative to the camera coordinate system, velocity $\mathbf{v}^c_{bk}$, direction $\mathbf{q}^c_{bk}$, accelerometer offset $\mathbf{b}_a$, and gyroscope offset $\mathbf{b}_g$. n is the total number of keyframes, and m is the total number of features in the sliding window. λ_i is the inverse distance of the ith feature from its first observation. $\mathbf{u}_h$ is the UWB localization result corresponding to the hth frame.

The maximum likelihood estimation problem can be framed as a nonlinear optimization issue. The residual component $\mathbf{E}_{VI}$ linked to visual and IMU measurements is expressed as:

$$\begin{aligned}\mathbf{E}_{VI} = &\left\|\mathbf{r}_p - \mathbf{H}_p\mathbf{X}_{VI}\right\|^2 + \sum_{k\in\beta}\left\|\mathbf{r}_\beta(\hat{\mathbf{z}}^{b_k}_{b_{k+1}}, \mathbf{X}_{VI})\right\|^2_{\mathbf{P}^{b_k}_{b_{k+1}}} \\ &+ \sum_{(l,j)\in\mathbf{C}} \rho\left(\left\|\mathbf{r}_c\left(\hat{\mathbf{z}}^{c_j}_l, \mathbf{X}_{VI}\right)\right\|^2_{\mathbf{P}^{c_j}_l}\right),\end{aligned} \tag{7}$$

where $\rho(\cdot)$ represents the Huber norm. The first term of the nonlinear least squares problem includes $\mathbf{r}_p$ and $\mathbf{H}_p$, which correspond to the prior information from marginalization. $\mathbf{r}_\beta(\hat{\mathbf{z}}^{b_k}_{b_{k+1}}, \mathbf{X}_{VI})$ and $\mathbf{r}_c(\hat{\mathbf{z}}^{c_j}_l, \mathbf{X}_{VI})$ are the residuals of inertial and visual, respectively. $\mathbf{P}^{b_k}_{b_{k+1}}$ and $\mathbf{P}^{c_j}_l$ are related measurement covariances, respectively.

In order to better combine the camera, IMU and UWB sensors to reduce the drift problem during localization. We introduce the residual term $\mathbf{E}_U$ of the UWB sensor in the optimization process. Then, the overall cost function $\mathbf{E}$ of the whole system is given as

$$\mathbf{E} = \mathbf{E}_{VI} + \mathbf{E}_U. \tag{8}$$

Unlike the residual term of UWB in other methods, we use a new residual function to optimize the overall state. Specifically, we calculate the Euclidean distance of UWB localization results on

adjacent VIO timestamps, formulated as follows:

$$\mathbf{E}_U = \sum \left\| \mathbf{T}_C^G \mathbf{p}_k^C - \mathbf{p}_k^G \right\|_{\mathbf{P}_U}^2. \tag{9}$$

At time step k, $\mathbf{p}_k^C$ and $\mathbf{p}_k^G$ represent the UAV positions in the VIO coordinate system and the global coordinate system, respectively. The transformation from the VIO coordinate system to the global coordinate system is denoted by $\mathbf{T}_C^G$, which is necessary for calculating the positions in the two coordinate systems. The covariance matrix of each UWB measurement position is represented by $\mathbf{P}_U$.

The Levenberg-Marquardt method is employed to estimate the state at each moment in the aforementioned cost function. To improve the efficiency and accuracy of our system, we utilize the sliding window and marginalization method. To solve this nonlinear problem, we adopt the Ceres solver in this article.

5.2 Loosely Coupled Stage

As previously discussed, integrating UWB localization results as a translation constraint in VIO has effectively reduced accumulated drift. However, the relative localization information obtained from this integration does not directly yield knowledge of the global position of UAV. We employ particle filter to obtain precise global localization information, which combines UWB localization results with the relative pose estimation derived from the tightly coupled stage. In the following sections, we will analyze by outlining the steps implemented in the particle filtering process:

Initialize. The loosely coupled stage receives input from VIO results $\{\mathbf{t}_k, < \mathbf{q}_{t_k}, \mathbf{t}_{t_k} >\}$, and UWB results $\{\mathbf{t}_k, < \mathbf{x}_{t_k}, \mathbf{y}_{t_k}, \mathbf{z}_{t_k} >\}$, where $\mathbf{t}_k$ is the timestamp while the two subsystems have been aligned. Our system combines these results in particle filter to obtain UAV global localization. We denote the set of particles as $\mathbf{P} = \{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_n\}$, where the state $\mathbf{P}_i =< \mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i, \mathbf{q}_i, \rho_i >$ of each particle consists of position $(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i)$, attitude $\mathbf{q}_i$ and weight ρ_i. During initialization, particles are assigned the same weight i.e., $\rho_i = \frac{1}{n}, i = 1, 2, \ldots, n$.

Predict. In the prediction step, we will predict the state of each particle based on the pose transformation and the position estimation. However, obtaining an accurate pose transformation matrix is one of the most critical techniques when using the tightly coupled stage to predict the current stage. After aligning UWB sensor timestamps with the timestamps of the tightly coupled phase, we find that there may be multiple pose transformations within the timestamp interval of the tightly coupled phase. To solve this problem, we need to predict the states of particles by achieving the pose transformation between two timestamps.

More specifically, we have aligned the timestamps $\mathbf{t}_k^G$ and $\mathbf{t}_{k+1}^G$ of UWB sensor with the timestamps $\mathbf{t}_k^L$ and $\mathbf{t}_{k+1}^L$ of tightly coupled stage, respectively. If there are no additional timestamps within the $\mathbf{t}_k^L$ and $\mathbf{t}_{k+1}^L$ timestamps, we construct the transformation matrix $\mathbf{T}_k^{k+1}$ by utilizing rotation quaternions and translation vectors from $\mathbf{t}_{k+1}^L$, facilitating the conversion between $\mathbf{t}_k^G$ and $\mathbf{t}_{k+1}^G$:

$$\mathbf{T}_k^{k+1} = \begin{pmatrix} \mathbf{R}_q & \mathbf{t} \\ \mathbf{0}^T & \mathbf{1} \end{pmatrix}, \tag{10}$$

where the rotation matrix $\mathbf{R}_q$ is a transformation from $\mathbf{t}_k^L$ to $\mathbf{t}_{k+1}^L$. This matrix is derived from the quaternion $\mathbf{q} = (\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z})$ within $\mathbf{t}_{k+1}^L$, and its components are given by:

$$\mathbf{R}_q = \begin{pmatrix} 1 - 2\mathbf{y}^2 - 2\mathbf{z}^2 & 2\mathbf{xy} - 2\mathbf{wz} & 2\mathbf{xz} - 2\mathbf{wy} \\ 2\mathbf{xy} + 2\mathbf{wz} & 1 - 2\mathbf{x}^2 - 2\mathbf{z}^2 & 2\mathbf{yz} - 2\mathbf{wx} \\ 2\mathbf{xz} - 2\mathbf{wy} & 2\mathbf{yz} + 2\mathbf{wx} & 1 - 2\mathbf{x}^2 - 2\mathbf{y}^2 \end{pmatrix}. \tag{11}$$

If there are additional timestamps within timestamps $\mathbf{t}_k^L$ and $\mathbf{t}_{k+1}^L$, assume that there are $S-1$ timestamps, for example, $\mathbf{t}_k^L, \mathbf{t}_{k,1}^L, \ldots, \mathbf{t}_{k,S-1}^L, \mathbf{t}_{k+1}^L$, then we let the transformation matrix in the time interval $[\mathbf{t}_{k,m}^L, \mathbf{t}_{k,m+1}^L]$ be $\mathbf{T}_{k,m}^{k,m+1}$. Based on the multiplication of transformation matrices, we can calculate the complete transformation matrix $\mathbf{T}_k^{k+1}$ in the time $[\mathbf{t}_k^L, \mathbf{t}_{k+1}^L]$ is given as follows:

$$\mathbf{T}_k^{k+1} = \mathbf{T}_{k,S-1}^{k+1}\mathbf{T}_{k,S-2}^{k,S-1}\cdots\mathbf{T}_{k,1}^{k,2}\mathbf{T}_k^{k,1}. \tag{12}$$

With the above calculation, we compute the tightly coupled stage complete transformation matrix $\mathbf{T}_k^{k+1}$ during the $[\mathbf{t}_k^L, \mathbf{t}_{k+1}^L]$ time interval. The 3D position $\mathbf{p}_{k,j}^G = (\mathbf{x}_{k,j}, \mathbf{y}_{k,j}, \mathbf{z}_{k,j})^{\mathrm{T}}$ of the sampled particle population at time $\mathbf{t}_k^G$ is written in the homogeneous form $\tilde{\mathbf{p}}_{k,j}^G = (\mathbf{x}_{k,j}, \mathbf{y}_{k,j}, \mathbf{z}_{k,j}, \mathbf{1})^{\mathrm{T}}$. The homogeneous coordinates of the predicted position of each particle at time $\mathbf{t}_{k+1}^G$ are calculated from the transformation matrix:

$$\tilde{\mathbf{p}}_{k+1,j}^G = \mathbf{T}_k^{k+1}\tilde{\mathbf{p}}_{k,j}^G, j = 0, 1, 2, \ldots, N. \tag{13}$$

Therefore, we can obtain predicted position $\mathbf{p}_{k+1,j}^G = (\mathbf{x}_{k+1,j}, \mathbf{y}_{k+1,j}, \mathbf{z}_{k+1,j})^{\mathrm{T}}$ of each particle at time $\mathbf{t}_{k+1}^G$.

Update. In the update step, each particle's weight is updated by using UWB observation data, leading to a more accurate posterior probability distribution. First, we obtain the UWB observation data $\mathbf{p}_{k+1}^G$ at the current time $\mathbf{t}_{k+1}^G$, which provides the target's position information. Next, we calculate the importance weights for each particle. This approach involves evaluating the distance $\mathbf{d}_{k+1,j}$ between the predicted state $\mathbf{p}_{k+1,j}^G$ of each particle and the actual observed data $\mathbf{p}_{k+1}^G$ and converting the difference into a critical weight. The distance is defined as:

$$\mathbf{d}_{k+1,j} = \left\| \mathbf{p}_{k+1}^G - \mathbf{p}_{k+1,j}^G \right\|. \tag{14}$$

Specifically, the weight $\rho_{k+1,j}$ for each particle is established by utilizing the distance from the observed data. Particles that are closer to UWB detection results will have higher weights, which are additionally updated as:

$$\rho_{k+1,j} = 1/\mathbf{d}_{k+1,j}. \tag{15}$$

Resample. In resampling step, a new set of particles is chosen based on their weights to better represent the posterior probability distribution. We perform the following operations:

(1) Normalize the weight of particles

$$\bar{\rho}_{k+1,j} = \frac{\rho_{k+1,j}}{\sum_{j=1}^{m}\rho_{k+1,j}}, \tag{16}$$

where m is the current number of particles.

(2) Generate a sequence of random numbers between $[0, 1)$. These random numbers will be used to sample from the collection of particles to decide which particles need to be replicated.

(3) Resample particles within an ellipse:

$$|(\mathbf{x}_n, \mathbf{y}_n, \mathbf{z}_n) - \mathbf{L}_{VI}| + |(\mathbf{x}_n, \mathbf{y}_n, \mathbf{z}_n) - \mathbf{L}_U| \leq 2\mathbf{D}, \tag{17}$$

where $\mathbf{L}_{VI}$ and $\mathbf{L}_U$ are VIO and UWB location results, respectively; and $\mathbf{D}$ is the maximum positioning error.

(4) Generate a particle with index n, denoted as:

$$\mathbf{X}_n =< \mathbf{x}_n, \mathbf{y}_n, \mathbf{z}_n, \mathbf{h}_n, \rho_n > . \tag{18}$$

(5) Repeat steps (2), (3), and (4) until the desired number of resampling particles is reached.

Fig. 4. Hardware setup and experimental environment.

In the resampling process, these steps improve the approximation of particle set to the target distribution, eliminating low-weight particles, and enhancing the filter performance and robustness.

Estimate. Due to the fluctuation of VIO position and UWB position in indoor environments during particle movement, we chose to calculate the weighted average of all particles as the target position:

$$(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \sum_{n=1}^{N} (\mathbf{x}_n, \mathbf{y}_n, \mathbf{z}_n)\, \rho_n, \tag{19}$$

which can make the obtained position estimate more robust.

6 EXPERIMENT

In this section, we present three sets of experiments assessing our proposed method. The first set entails executing the D-RANSAC algorithm across diverse indoor motion trajectories to gauge its efficacy. The second set involves comprehensive benchmark tests using the EuRoC public dataset on the HCCNet system, where we compare our approach against leading VIO methods to underscore its accuracy and robustness. Lastly, the third set implements the HCCNet system within complex indoor environments to assess its positioning performance.

Throughout these experiments, data collection employs the Intel RealSense Depth Camera D435i and the Decawave UWB RF Module DWM1000. The D435i camera, a structured light depth camera, provides monocular image content, while its integrated IMU enhances localization accuracy. The DWM1000 module employs UWB technology for precise ranging and localization capabilities, measuring UAV-to-base station distances via signal time delay and multipath propagation effects. As shown in Figure 4, both components are accurately calibrated to align acquired data with the environment. The backend optimization module runs on an Intel(R) Celeron(R) G4900T CPU @ 2.90 GHz, 16.0 GB RAM server operating Ubuntu 16.04.

6.1 D-RANSAC Algorithm for Outlier Filtering

In this experiment, we focus on verifying the effectiveness of the proposed D-RANSAC algorithm in detecting and rejecting outliers in UWB localization results under indoor environmental interference.

Experimental Setup. We conduct experiments in representative indoor scenes, including classroom buildings and office buildings, based on the outliers generated by sensors and different

Table 1. Comparison of UWB Localization Mean Error (ME) Performance under Different Methods

ME \ Method / Dataset	LS ($n = 4$)	LS ($n = 5$)	LS ($n = 6$)	EKF ($n = 4$)	UKF ($n = 4$)	TS-KF ($n = 4$)	SL-LS ($n = 4$)	D-RANSAC ($n = 4$)
O/Straight-Line	0.1128	0.1199	0.1004	0.1067	0.0998	0.0973	0.1074	**0.0963**
O/L-shaped	0.1216	0.1096	0.0926	0.1048	0.0974	0.1157	0.1001	**0.0910**
O/U-shaped	0.1115	0.1041	0.1046	0.1067	0.1022	0.0988	0.1069	**0.0986**
O/Z-shaped	0.1133	0.1080	0.1022	0.0996	0.1154	0.1071	0.1006	**0.0926**
C/Straight-Line	0.1074	0.0966	0.0965	0.1031	0.0945	0.1034	0.1129	**0.0900**
C/L-shaped	0.1059	0.0950	**0.0874**	0.0917	0.0928	0.0941	0.0909	0.0885
C/U-shaped	0.1035	0.0918	**0.0902**	0.0972	0.0931	0.0920	0.0964	0.0904
C/Z-shaped	0.1049	0.0831	0.0843	0.0914	0.1121	0.0856	0.0901	**0.0826**

[1] The values in bold are the smallest values in their respective rows.
[2] "O" represents office scenario, while "C" denotes classroom scenario.
[3] The letter "n" indicates the number of UWB anchors.

environmental conditions. These scenes have different spatial geometries and environmental dynamics, providing different challenges for estimating localization results. In the classroom scene, there is only minimal interference from radio signals, and the geometric shapes of the scene are regular, with objects in close proximity to the sensing devices. However, in the office scene, many interfering radio signals increase the likelihood of UWB outliers. We collect trajectory data specific to different scenes in our experiments. Fixed motion trajectories of UWB tags are set, including straight-line, L-shaped, U-shaped, and Z-shaped trajectories. We simulate outliers by temporarily occluding the tag during its movement.

Quantitative Comparison. We assess the localization algorithm primarily by utilising of the mean error (ME) as a fundamental metric. Moreover, we examine the stability and precision of the outcomes by calculating the standard deviation (SD) associated with the localization data. To ensure a comprehensive evaluation, we introduce an additional metric, the maximum absolute error (MAX), into our analysis.

We conduct comprehensive experiments to assess the D-RANSAC algorithm thoroughly. Our evaluation includes various numbers of UWB anchors, motion trajectories, and comparisons with established indoor localization techniques in different settings. We benchmark D-RANSAC against EKF, **Unscented Kalman Filter (UKF)**, and Taylor series-based location estimation technique (TS-KF) [37].

The EKF approximates nonlinear functions in the Kalman filter using a first-order Taylor series. It's known for simplicity and handling small to moderate nonlinearities well. The UKF, an EKF improvement, uses sigma points to better capture the state distribution. This method excels under strong nonlinearities. The TS-KF, a recent development, uses the Taylor series for direct location estimation. It's efficient and accurate, especially with rapid motion changes.

We present experimental outcomes, including ME, SD, and MAX for each dataset, in Tables 1–3. These tables compare D-RANSAC with EKF, UKF, and TS-KF. This comparison shows D-RANSAC's strengths. It is robust to environmental changes and accurate with few UWB anchors.

In the context of the office experimental scene, employing four UWB anchors, the D-RANSAC algorithm demonstrates substantial enhancements in localization accuracy across different trajectory shapes: straight-line, L-shaped, U-shaped, and Z-shaped. Specifically, the average localization errors for these trajectories are 0.0963 m, 0.0910 m, 0.0986 m, and 0.0926m, respectively. These improvements amount to 14.63%, 25.22%, 11.57%, and 18.34% compared to the original

Table 2. Comparison of UWB Localization Standard Deviation (SD) Performance under Different Methods

SD / Method / Dataset	LS ($n = 4$)	LS ($n = 5$)	LS ($n = 6$)	EKF ($n = 4$)	UKF ($n = 4$)	TS-KF ($n = 4$)	SL-LS ($n = 4$)	D-RANSAC ($n = 4$)
O/Straight-Line	0.0758	0.0893	0.0685	0.1590	0.1084	0.0595	0.0729	**0.0586**
O/L-shaped	0.0887	0.0813	0.0731	0.0914	0.0728	0.0690	0.0772	**0.0654**
O/U-shaped	0.0776	0.0737	0.0807	0.1938	0.1305	0.0787	0.0776	**0.0769**
O/Z-shaped	0.0898	0.0835	0.0797	0.1471	0.0875	0.1340	0.0770	**0.0750**
C/Straight-Line	0.0857	0.0645	0.0679	0.1577	0.1090	0.0684	0.0869	**0.0677**
C/L-shaped	0.0779	0.0791	0.0663	0.1031	0.0709	0.0698	0.0751	**0.0659**
C/U-shaped	0.0799	0.0711	0.0684	0.2033	0.1364	0.0700	0.0821	**0.0676**
C/Z-shaped	0.0786	0.0666	0.0605	0.1468	0.0848	0.1217	0.0650	**0.0602**

[1] The values in bold are the smallest values in their respective rows.
[2] "O" represents office scenario, while "C" denotes classroom scenario.
[3] The letter "n" indicates the number of UWB anchors.

Table 3. Comparison of UWB Localization Maximum Absolute Error (MAX) Performance under Different Methods

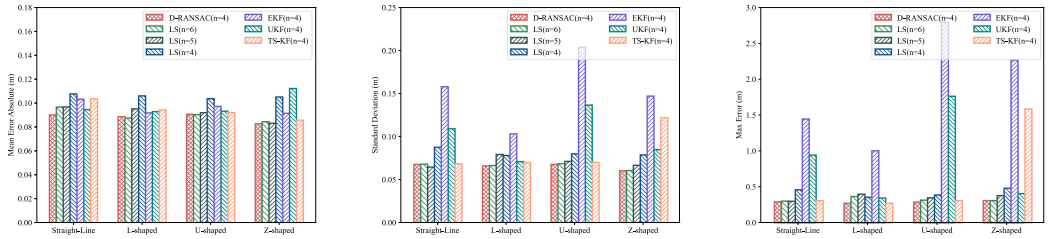
MAX / Method / Dataset	LS ($n = 4$)	LS ($n = 5$)	LS ($n = 6$)	EKF ($n = 4$)	UKF ($n = 4$)	TS-KF ($n = 4$)	SL-LS ($n = 4$)	D-RANSAC ($n = 4$)
O/Straight-Line	0.3467	0.4114	0.2853	1.4969	0.9954	0.2510	0.3467	**0.2487**
O/L-shaped	0.3457	0.4206	0.3679	0.7118	0.3643	0.2991	0.3457	**0.2463**
O/U-shaped	0.3796	0.3753	0.3590	2.6518	1.6820	0.5177	0.3796	**0.3111**
O/Z-shaped	0.4762	0.3977	0.3578	2.2073	0.4443	1.777	**0.3413**	0.3579
C/Straight-Line	0.4591	0.2997	0.3010	1.4438	0.9417	0.3068	0.4519	**0.2915**
C/L-shaped	0.3546	0.3976	0.3654	1.0012	0.3439	0.2726	0.3546	**0.2712**
C/U-shaped	0.3868	0.3474	0.3174	2.7877	1.7600	0.3074	0.3764	**0.2876**
C/Z-shaped	0.4800	0.3779	**0.3085**	2.2628	0.4061	1.5822	0.4800	0.3090

[1] The values in bold are the smallest values in their respective rows.
[2] "O" represents office scenario, while "C" denotes classroom scenario.
[3] The letter 'n" indicates the number of UWB anchors.

localization outcomes. Similarly, the algorithm reduces localization standard deviations by 22.69%, 26.27%, 0.91%, and 16.36%, as depicted in Figures 5(a) and 5(b).

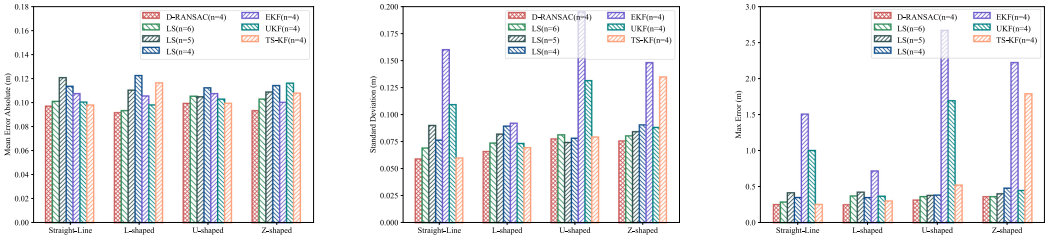
In the classroom building scene, employing four UWB anchors, the D-RANSAC algorithm also demonstrates notable accuracy enhancements. Average localization errors of 0.0900 m, 0.0885m, 0.0904m, and 0.0826m are observed for the same trajectory shapes, translating to accuracy improvements of 16.20%, 16.43%, 12.66%, and 21.26% compared to the original localization results. Correspondingly, standard deviations reduce by 22.63%, 15.40%, 15.39%, and 23.41%, as illustrated in Figures 6(a) and 6(b).

In testing L-shaped and U-shaped paths inside classroom scenes, the D-RANSAC algorithm often shows suboptimal performance. This issue likely stems from too few anchor points. This lack makes UWB localization very sensitive to the environment along these specific paths. Such situations worsen mean errors due to multipath effects and **non-line-of-sight (NLOS)** propagation. This impacts D-RANSAC's overall performance. However, increasing anchor points to six



(a) Mean Error (b) Standard Deviation (c) Maximum Error

Fig. 5. Comparison in classroom environment.



(a) Mean Error (b) Standard Deviation (c) Maximum Error

Fig. 6. Comparison in office environment.

improves the least squares method's resilience to these issues. It then performs better in terms of average error. This highlights the need to consider environmental factors and anchor placement strategies in designing indoor localization systems.

Moreover, when applying the D-RANSAC algorithm to the original localization outcomes with four UWB anchors, the results closely parallel those achieved with six UWB anchors. This indicates that the algorithm is effective in enhancing localization accuracy and mitigating environmental interference, particularly in complex real-world scenarios.

Furthermore, our investigation of the maximum impact of outliers in indoor environments (Figures 5(c) and 6(c)) highlights the algorithm's significant reduction of maximum positioning errors in both the classroom and office building environments. This robustness underscores the D-RANSAC algorithm's ability to handle indoor interference challenges.

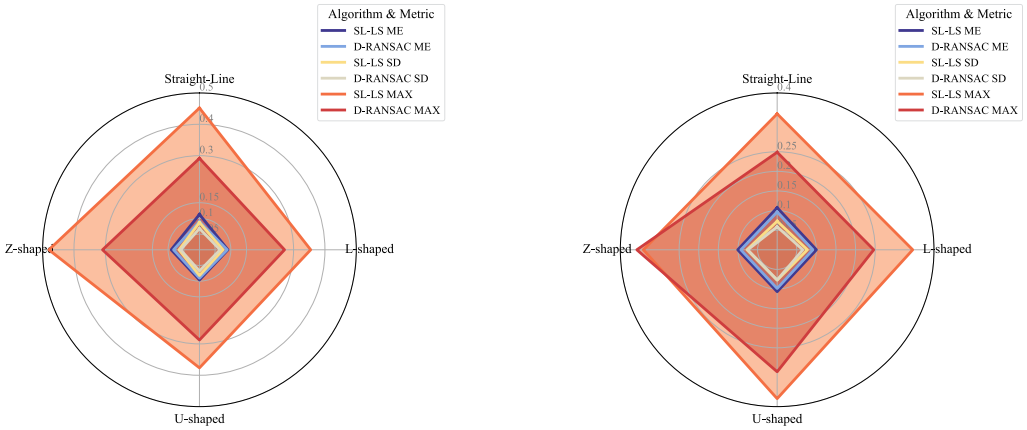
Statistical Advantages of D-RANSAC. To further highlight the D-RANSAC algorithm's robustness and efficacy in indoor localization, we conduct performance comparisons. Using the Mann-Whitney U test, we assess differences between D-RANSAC and other methods, including LS, EKF, UKF, and TS-KF. Our evaluation covers two indoor environments: office and classroom. Through these comparisons, we aim to show D-RANSAC's advantages in indoor localization tasks. We use P-values from these tests to check if differences in localization accuracy among the methods are significant. We set the significance level threshold at 0.05, as Table 4 shows.

Our study currently shows that the D-RANSAC algorithm outshines conventional methods significantly. In office settings, we notice P-values dropping far below the 0.05 mark. This marks a notable statistical edge of D-RANSAC over other approaches. Classrooms reveal a similar trend, affirming D-RANSAC's utility in diverse indoor situations. Yet, for specific paths in certain places, the P-values don't highlight significance, indicating room for enhancement. Overall, our data convincingly prove D-RANSAC significantly boosts indoor localization performance.

Table 4. Mann-Whitney U Test P-Values Comparing D-RANSAC with Other Localization Methods

P Value / Method / Dataset	LS ($n = 4$)	LS ($n = 5$)	LS ($n = 6$)	EKF ($n = 4$)	UKF ($n = 4$)	TS-KF ($n = 4$)
Office	2.42×10^{-4}	3.16×10^{-2}	6.32×10^{-2}	4.87×10^{-2}	7.97×10^{-3}	8.17×10^{-24}
Classroom	2.10×10^{-7}	2.29×10^{-2}	2.81×10^{-2}	1.62×10^{-4}	3.96×10^{-7}	1.45×10^{-22}

[1] The letter "n" n indicates the number of UWB anchors.



(a) Comparison of Different Metrics in Office.

(b) Comparison of Different Metrics in Classroom.

Fig. 7. Ablation experiments in different environments.

Ablation Study. In this study, we conduct a series of ablation experiments. Our aim is to see how our D-RANSAC algorithm stacks up against traditional methods. This focus is on filtering outliers in complex trajectory data. Due to RANSAC's limits–mainly its fit for straight-line paths–we use the sliding window least squares (SL-LS) method for comparison. This helps us check its efficiency in spotting and managing outliers. Our analysis zeroes in on three main performance metrics: ME, SD, and MAX in localization, as Tables 1–3 show.

Figures 7(a) and 7(b) visually show the measurement comparison in office and classroom environment, respectively. The results are consistent and highlight D-RANSAC's superior performance in all metrics. Specifically, for ME, D-RANSAC significantly reduces errors, showing its precision. In evaluating SD, D-RANSAC's lower values prove its consistent localization results. Also, D-RANSAC maintains the smallest range in MAX, showing its strength against extreme outliers.

These findings strongly support our hypothesis. By integrating the robustness of RANSAC with the flexibility of sliding window least squares, D-RANSAC not only overcomes traditional RANSAC's limitations in handling complex trajectories. It also significantly improves accuracy and reliability in filtering outliers. Thus, our research offers an effective solution for managing complex trajectory data full of outliers. It expands RANSAC's applicability across diverse scenarios.

Parameters Sensitivity Analysis. We perform parameter sensitivity analysis for the D-RANSAC algorithm's settings. These experiments aim to assess how different sliding window sizes ($M = 15, 25, 30, 35, 45$) and distance thresholds ($d = 5$cm, 8cm, 10cm, 13cm, 15cm) affect its performance. We measure these settings' impact on performance by comparing ME. The results appear in Tables 5 and 6.

Table 5. Comparison of D-RANSAC Algorithm's Maean Error Values under Various Sliding Window Size

ME \ Method / Dataset	D-RANSAC ($M = 15$)	D-RANSAC ($M = 25$)	D-RANSAC ($M = 30$)	D-RANSAC ($M = 35$)	D-RANSAC ($M = 45$)
O/Straight-Line	0.1245	0.1224	**0.0963**	0.0976	0.1144
O/L-shaped	0.1342	0.1241	**0.0910**	0.1205	0.0974
O/U-shaped	0.1193	0.1198	**0.0986**	0.1125	0.1222
O/Z-shaped	0.1125	0.1174	**0.0926**	0.1128	0.0098
C/Straight-Line	0.1131	0.1136	**0.0900**	0.1007	0.0945
C/L-shaped	0.1075	0.1059	**0.0885**	0.1121	0.1144
C/U-shaped	0.1097	0.0978	**0.0904**	0.1036	0.1164
C/Z-shaped	0.1009	0.0994	**0.0826**	0.1021	0.0964

[1] The values in bold are the smallest values in their respective rows.
[2] "O" represents office scenario, while "C" denotes classroom scenario.
[3] The letter "M" indicates the sliding window size.

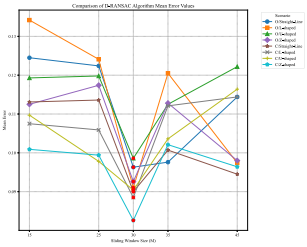
Table 6. Comparison of D-RANSAC Algorithm's Mean Error Values under Various Distance Threshold

ME \ Method / Dataset	D-RANSAC ($d = 5cm$)	D-RANSAC ($d = 8cm$)	D-RANSAC ($d = 10cm$)	D-RANSAC ($d = 13cm$)	D-RANSAC ($d = 15cm$)
O/Stright-Line	0.1163	0.1340	**0.0916**	0.1214	0.1214
O/L-shaped	0.0956	0.0951	**0.0891**	0.1050	0.1014
O/U-shaped	0.0993	0.0997	**0.0986**	0.1037	0.1073
O/Z-shaped	0.1154	0.1164	**0.0926**	0.1175	0.1093
C/Stright-Line	0.0928	0.0950	**0.0900**	0.0995	0.1051
C/L-shaped	0.0869	0.0929	**0.0885**	0.0979	0.0936
C/U-shaped	0.0925	0.0926	**0.0904**	0.0926	0.0920
C/Z-shaped	0.0853	0.0856	**0.0826**	0.0837	0.0830

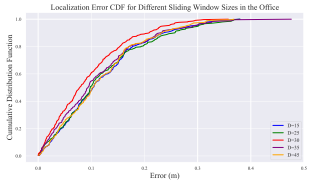
[1] The values in bold are the smallest values in their respective rows.
[2] "O" represents office scenario, while "C" denotes classroom scenario.
[3] The letter d indicates the distance threshold.

In Table 5, D-RANSAC achieves the smallest maximum error values for all trajectory types with a sliding window size of 30. Figure 8(a) gives a specific demonstration. Figures 8(b) and 8(c) respectively depict **cumulative distribution functions (CDFs)** of localization errors within office and classroom settings. This suggests that a sliding window size of 30 is ideal for optimal localization accuracy. This ideal size allows D-RANSAC to balance detail and overview. With 30 points, it captures enough trajectory patterns. It avoids noise that smaller windows might miss. Larger windows could dilute important details. Thus, size 30 offers the best accuracy for different paths. It suits the dynamic nature of indoor movements. This makes D-RANSAC versatile for various indoor scenarios.

Similarly, Table 6 shows a 10-cm distance threshold works best. Figure 9(a) gives a specific demonstration. Figures 9(b) and 9(c), respectively, depict CDFs of localization errors within office and classroom settings. It gives D-RANSAC the lowest errors in all paths, in both office and



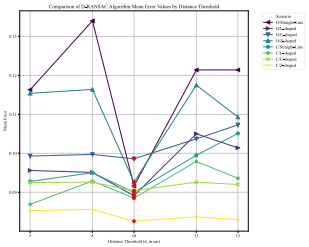
(a) Comparison of ME.



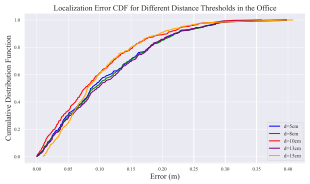
(b) Localization Error CDF in Office.

(c) Localization Error CDF in Classroom.

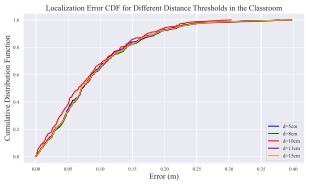
Fig. 8. Comparison of metrics for different sliding windows in different environments.



(a) Comparison of ME.



(b) Localization Error CDF in Office.



(c) Localization Error CDF in Classroom.

Fig. 9. Comparison of metrics for different distance thresholds in different environments.

classroom settings. This confirms our distance choice boosts precision well. A 10-cm threshold balances accuracy and flexibility. It's tight enough to filter out large errors but forgiving for minor inaccuracies. This setting helps D-RANSAC ignore minor fluctuations while catching real outliers. It suits the varied indoor dynamics, ensuring reliability across different areas. Thus, this threshold enhances D-RANSAC's ability to pinpoint locations accurately.

These experimental results back the practical use of our proposed parameters. They also show D-RANSAC's potential in indoor localization. Through these well-planned experiments, we believe the chosen parameters deliver accurate and reliable services.

6.2 Comparison on EuRoC Public Dataset

In this section, we compare the proposed HCCNet system with state-of-the-art VIO systems using a publicly available benchmark dataset.

Experimental Setup. The performance assessment of the HCCNet system takes place on the publicly accessible EuRoC dataset. This dataset comprises 11 visual-inertial datasets captured by a micro aerial vehicle (MAV) as it navigates through three distinct static scenes. These scenes include two VICON rooms (V1 and V2) and a machine hall (MH). However, since the EuRoC dataset lacks publicly available UWB sensor localization data, we make appropriate adjustments to simulate the input of UWB positioning. In our simulation experiments, we add Gaussian-distributed noise to the real data of the EuRoC dataset. This noise is configured to have deviations of 10 cm in the x and y dimensions, and 50 cm along the z-axis. By introducing this noise, we can simulate the uncertainty and errors of actual UWB positioning and provide it as input to our algorithm.

We conduct benchmark tests on HCCNet and compare it with top VIO systems. These include MSCKF [23], OKVIS [12], ROVIO [2], VINS-MONO [33], EQVIO [39], EKF [15], and a UWB-integrated VIO method [45]. We also look at VIUNet [10], which uses deep learning for feature

Table 7. Comparison of Different VIO Methods in the EuRoC Dataset

ATE	MSCKF	ROVIO	VINS MONO	OKVIS	EQVIO	Tightly Coupled	VIUNet	EKF	HCCNet-D	HCCNet
MH01	0.4346	0.2982	0.1731	0.2258	0.1214	0.1483	0.1046	0.1385	0.0911	**0.0593**
MH02	0.4389	0.3297	0.0813	0.1692	0.1726	0.0919	0.0997	0.0887	0.0947	**0.0636**
MH03	0.2517	0.3925	0.1501	0.2682	**0.1119**	0.1338	0.2501	0.1452	0.1427	0.1275
MH04	0.6108	0.7237	0.1882	0.2248	0.1721	0.1437	0.1748	0.1202	0.1466	**0.1146**
MH05	0.4853	1.0533	0.2603	0.3002	0.2521	0.1268	0.1621	0.1198	0.1373	**0.1123**
V1_01	0.2983	0.1561	0.0634	0.0857	0.0554	0.0909	–	0.0672	0.0880	**0.0527**
V1_02	0.2011	0.1938	0.2591	0.1801	0.1344	0.1015	0.0809	0.1253	0.1221	**0.0958**
V1_03	0.6744	0.1699	0.1583	0.1918	0.1863	0.1057	0.0953	0.1257	0.1151	**0.0904**
V2_01	0.1123	0.2566	0.0795	0.0829	0.0785	0.0868	0.0751	0.0802	0.0820	**0.0409**
V2_02	0.1652	0.5353	2.0897	0.1551	0.1646	2.0339	0.0848	0.1211	0.1126	**0.0848**
V2_03	1.1344	0.1882	0.1494	1.9879	0.1935	**0.1093**	0.0999	0.1319	0.1349	0.1199

[1] The values in bold are the smallest values in their respective rows.
[2] "–" denotes using V1_01 as the VIUNet model validation sequence.

extraction. It excels in complex indoor localization. We measure their performance on the EuRoC dataset by the **root mean square error (RMSE)** of the **absolute trajectory error (ATE)** in meters. For each dataset, we run 10 trials using EuRoC's default calibration. We then calculate the median RMSE ATE as the final result.

Results and Evaluation. The performance evaluation, presented in Table 7, underscores the successful operation of the proposed HCCNet system across all datasets, yielding strong results across the entire range of EuRoC datasets. Our system consistently outperform alternative methods, particularly evident in the context of ATE, where it achieves the lowest error values. This exceptional accuracy and precision highlight our system's superiority.

Interestingly, in comparing VIO methods on the EuRoC dataset, HCCNet now shows top performance. Table 7 reveals HCCNet leads over MSCKF, ROVIO, VINS-MONO, OKVIS, and EQVIO in ATE. The review also looks at VIUNet and EKF. VIUNet uses deep learning for features, doing well in complex indoor areas. Yet, it doesn't beat HCCNet. EKF performs well on some datasets but still has higher ATE than HCCNet. Importantly, HCCNet often scores the lowest ATE, underlining its edge in accuracy and stability.

Ablation Study. In evaluating VIO methods on the EuRoC dataset, we examine the performance difference between HCCNet-D and the full HCCNet. HCCNet-D is a version of HCCNet without the D-RANSAC algorithm. By comparing their ATE, we assess D-RANSAC's impact on HCCNet's performance.

Table 7 shows HCCNet consistently has lower ATE values than HCCNet-D. This indicates that D-RANSAC greatly improves localization accuracy. Especially in datasets like MH01, MH02, V1_01, and V2_01, HCCNet's ATE values are much lower than HCCNet-D's. This highlights D-RANSAC's crucial role in outlier management and localization robustness. D-RANSAC optimizes localization by accurately identifying and removing outliers, thus achieving higher precision in complex environments. These findings underline D-RANSAC's importance in HCCNet, enhancing both accuracy and system stability in tough environments. Therefore, D-RANSAC is vital to HCCNet, providing a more reliable and precise indoor localization solution.

From a broader perspective, our experimental results robustly establish the effectiveness and reliability of the proposed HCCNet system. They affirm substantial enhancements in localization accuracy and precision when contrastes with existing VIO methods. This accomplishment

Table 8. Comparison of Trajectory Running Data in Real Environment

ATE	Classroom		Hall		Office	
	Loop	Random	Loop	Random	Loop	Random
VINS-Mono	0.1429	0.5314	0.1377	0.5319	0.1441	0.5325
Tightly Coupled	0.1259	0.1291	0.1133	0.1013	0.1404	0.1444
HCCNet	**0.0934**	**0.0920**	**0.0924**	**0.0904**	**0.0931**	**0.0910**

[1] The values in bold are the smallest values in their respective columns.

underscores our system's considerable contribution to the field, promising advanced solutions for accurate and reliable localization tasks.

6.3 Comparison in Real Environment

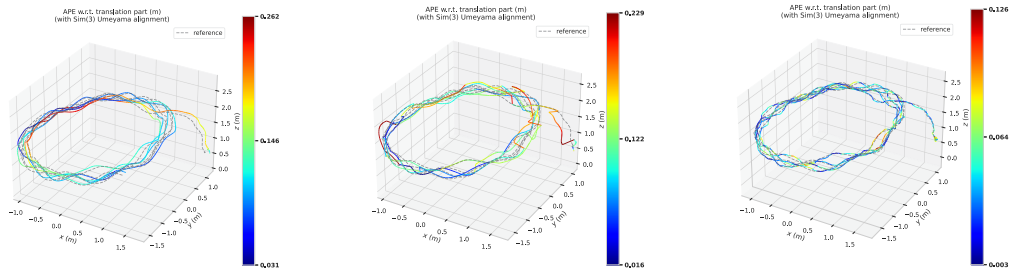
In this section, we further evaluate the performance of the proposed HCCNet system in real indoor environments.

Experimental Setup. Our objective differ from outlier detection and removal experiments, so we collect a specific dataset for testing the HCCNet system in indoor environments. We establish different benchmark datasets based on different scene environments, considering major factors such as environmental interference and information from the capturing devices in real-world scenarios. We record two types of trajectories in three different environments.

We use a UAV with an Intel RealSense Depth Camera D435i and a Decawave UWB RF module DWM1000 to collect benchmark datasets in classroom buildings, offices, and halls. To secure high-quality data, we guide the mobile robot to move at a slow pace along predetermined paths in distinct datasets during the data acquisition phase. We use the AprilTag method [30] to obtain ground truth data for the experiment. AprilTag is an image-based target recognition and localization system that identifies and measures tags with specific patterns. These tags typically appear as 2D codes with unique encoding and geometric features that can be accurately detected and located by the camera in the image. We use pre-calibrated AprilTag markers and place them at key positions in the experimental scenes. The camera captures the positions and orientations of these tags in real time by observing them in the scene. In each time step of the experiment, we process and recognize the AprilTag markers in the images to obtain the positions and orientations of the tags in the real world. These data is considered as ground truth for comparison and evaluation with the estimation results of our proposed localization algorithm.

Results and Evaluation. As shown in Table 8, HCCNet system successfully run on all datasets in various indoor environments and obtained more accurate results in terms of RMSE ATE compared to other benchmarked systems. A comparison is made between VINS-Mono and our proposed system. For methods that consider only visual-inertial information, accumulated drift may not be corrected over time, leading to more pronounced errors. By incorporating UWB data and visual-inertial information as constraints, the system is improved, further enhancing the system's performance by more in-depth particle filter fusion of UWB global localization results with the preprocessed final results.

To visualise the advantages of our system over others, we present in Figure 10 the trajectory error curves estimated by VINS-Mono (left) , tightly coupled stage (middle) and our system (right). Specifically, the positioning accuracy of VINS-Mono is significantly degraded over time, as it is based on a purely visual-inertial sensor and is subject to positioning drift. In contrast, our system consistently achieves ideal positioning results. This high accuracy and drift-free performance is



(a) VINS-MONO (b) Tightly Coupled Stage (c) HCCNet

Fig. 10. Trajectory error curves estimated by VINS-Mono (left), tightly coupled stage (middle) and our system (right) in classroom.

due to the combination with UWB sensor distance measurements, resulting in accurate positioning results.

These findings suggest the HCCNet system can provide high localization accuracy in different environmental conditions. It effectively reduces localization errors, resulting in a higher probability of positions with smaller errors. This system is significant for indoor localization applications, particularly in high-precision and reliability scenarios.

7 CONCLUSION

In this article, we proposed an innovative hybrid coupled fusion network to achieve accurate global localization estimation for UAVs. The localization results of UWB are detected and refined by the proposed D-RANSAC algorithm, and the optimized results are incorporated into the tightly coupled and the loosely coupled stages, respectively. We extensively tested our proposed HCCNet system on public datasets and in real-world experiments, and the results show that its localization accuracy is comparable to the state-of-the-art VIO systems. More importantly, we observed a significant reduction in drift accumulation during the experiment.

However, the proposed HCCNet system requires multiple UWB anchors to achieve localization and provides global localization results for a single UAV. Future research directions will mainly focus on two aspects: one is to reduce the number of UWB anchors required to reduce the complexity and cost of the system while maintaining localization accuracy, and the other is to achieve global localization results for multiple UAVs.

In the future, we plan to expand our work. First, we will explore additional parameter settings to further enhance positioning accuracy. Second, comparing D-RANSAC with other localization techniques, particularly in terms of cost and complexity, is also on our agenda. Furthermore, we aim to investigate the performance of HCCNet in various environments. Finally, with the potential for real-time applications in mind, enhancing algorithm efficiency is another goal. We anticipate that these efforts could lead to new breakthroughs in indoor positioning technology.

REFERENCES

[1] Muhammad Atif, Rizwan Ahmad, Waqas Ahmad, Liang Zhao, and Joel J. P. C. Rodrigues. 2021. UAV-assisted wireless localization for search and rescue. *IEEE Syst. J.* 15, 3 (Feb. 2021), 3261–3272.

[2] Michael Bloesch, Sammy Omari, Marco Hutter, and Roland Siegwart. 2015. Robust visual inertial odometry using a direct EKF-based approach. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS'15)*. 298–304.

[3] Stefano Bottigliero, Daniele Milanesio, Maurice Saccani, and Riccardo Maggiora. 2021. A low-cost indoor real-time locating system based on TDOA estimation of UWB pulse sequences. *IEEE Trans. Instrum. Meas.* 70 (2021), 1–11.

[4] Yanjun Cao and Giovanni Beltrame. 2021. VIR-SLAM: Visual, inertial, and ranging SLAM for single and multi-robot systems. *Auto. Robots* 45, 6 (Sep. 2021), 905–917.
[5] Hong-Beom Choi, Keun-Woo Lim, and Young-Bae Ko. 2023. LUVI: Lightweight UWB-VIO based relative positioning for AR-IoT applications. *Ad Hoc Netw.* 145 (2023), 103132.
[6] Martin A. Fischler and Robert C. Bolles. 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 6 (Jun. 1981), 381–395.
[7] Bo Gao, Baowang Lian, Dongjia Wang, and Chengkai Tang. 2022. Low drift visual inertial odometry with UWB aided for indoor localization. *IET Commun.* 16, 10 (Jun. 2022), 1083–1093.
[8] Zheng Gong, Peilin Liu, Fei Wen, Rendong Ying, Xingwu Ji, Ruihang Miao, and Wuyang Xue. 2021. Graph-based adaptive fusion of GNSS and VIO under intermittent GNSS-degraded environment. *IEEE Trans. Instrum. Meas.* 70 (2021), 1–16.
[9] KwangYik Jung, SungJae Shin, and Hyun Myung. 2022. U-VIO: Tightly coupled UWB visual inertial odometry for robust localization. In *Proceedings of the International Conference on Robot Intelligence Technology and Applications (RiTA'22)*, Vol. 429. Springer, 272–283.
[10] Peng-Yuan Kao, Hsiu-Jui Chang, Kuan-Wei Tseng, Timothy Chen, He-Lin Luo, and Yi-Ping Hung. 2023. VIUNet: Deep visual–inertial–UWB fusion for indoor UAV localization. *IEEE Access* 11 (2023), 61525–61534.
[11] Fabrizio Lazzari, Alice Buffi, Paolo Nepa, and Sandro Lazzari. 2017. Numerical investigation of an UWB localization technique for unmanned aerial vehicles in outdoor scenarios. *IEEE Sensors J.* 17, 9 (May 2017), 2896–2903.
[12] Stefan Leutenegger, Paul Furgale, Vincent Rabaud, Margarita Chli, Kurt Konolige, and Roland Siegwart. 2013. Keyframe-based visual-inertial SLAM using nonlinear optimization. In *Proc. Robot., Sci. Syst.*
[13] Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart, and Paul Furgale. 2015. Keyframe-based visual-inertial odometry using nonlinear optimization. *Int. J. Robot. Res.* 34, 3 (2015), 314–334.
[14] Danyang Li, Jingao Xu, Zheng Yang, Chenshu Wu, Jianbo Li, and Nicholas D Lane. 2021. Wireless localization with spatial-temporal robust fingerprints. *ACM Trans. Sensor Netw.* 18, 1 (2021), 1–23.
[15] Huei-Yung Lin and Jia-Rong Zhan. 2023. GNSS-denied UAV indoor navigation with UWB incorporated visual inertial odometry. *Measurement* 206 (2023), 112256.
[16] Kangcheng Liu and Ben M. Chen. 2023. Industrial UAV-based unsupervised domain adaptive crack recognitions: From database towards real-site infrastructural inspections. *IEEE Trans. Ind. Electron.* 70, 9 (Sep. 2023), 9410–9420.
[17] Ran Liu, Yongping He, Chau Yuen, Billy Pik Lik Lau, Rashid Ali, Wenpeng Fu, and Zhiqiang Cao. 2021. Cost-effective mapping of mobile robot based on the fusion of UWB and short-range 2-D LiDAR. *IEEE/ASME Trans. Mechatronics* 27, 3 (Jun. 2021), 1321–1331.
[18] Ran Liu, Chau Yuen, Tri-Nhut Do, Meng Zhang, Yong Liang Guan, and U-Xuan Tan. 2020. Cooperative positioning for emergency responders using self IMU and peer-to-peer radios measurements. *Inf. Fusion* 56 (Apr. 2020), 93–102.
[19] Tianxia Liu, Bofeng Li, Guang'e Chen, Ling Yang, Jing Qiao, and Wu Chen. 2024. Tightly coupled integration of GNSS/UWB/VIO for reliable and seamless positioning. *IEEE Trans. Intell. Transp. Syst.* 25, 2 (2024), 2116–2128.
[20] Valerio Magnago, Pablo Corbalán, Gian Pietro Picco, Luigi Palopoli, and Daniele Fontanelli. 2019. Robot localization via odometry-assisted ultra-wideband ranging with stochastic guarantees. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots Systems (IROS'19)*. 1607–1613.
[21] Ruben Mascaro, Lucas Teixeira, Timo Hinzmann, Roland Siegwart, and Margarita Chli. 2018. GOMSF: Graph-optimization based multi-sensor fusion for robust UAV pose estimation. In *Proceedings of the International Conference on Robotics and Automation (ICRA'18)*. 1421–1428.
[22] James McCoy, Atul Rawal, Danda B. Rawat, and Brian M. Sadler. 2022. Ensemble deep learning for sustainable multimodal UAV classification. *IEEE Trans. Intell. Transp. Syst.* (2022). doi:10.1109/TITS.2022.3170643.
[23] Anastasios I. Mourikis and Stergios I. Roumeliotis. 2007. A multi-state constraint Kalman filter for vision-aided inertial navigation. In *Proceedings of the International Conference on Robotics and Automation (ICRA'07)*. 3565–3572.
[24] Raul Mur-Artal and Juan D. Tardós. 2017. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Trans. Robot.* 33, 5 (Oct. 2017), 1255–1262.
[25] Raúl Mur-Artal and Juan D. Tardós. 2017. Visual-inertial monocular SLAM with map reuse. *IEEE Robot. Automat. Lett.* 2, 2 (Apr. 2017), 796–803.
[26] Thien Hoang Nguyen, Thien-Minh Nguyen, Muqing Cao, and Lihua Xie. 2020. Loosely-coupled ultra-wideband-aided scale correction for monocular visual odometry. *Unmanned Syst.* 8, 02 (2020), 179–190.
[27] Thien Hoang Nguyen, Thien-Minh Nguyen, and Lihua Xie. 2020. Tightly-coupled single-anchor ultra-wideband-aided monocular visual odometry system. In *Proceedings of the International Conference on Robotics and Automation (ICRA'20)*. 665–671.
[28] Thien Hoang Nguyen, Thien-Minh Nguyen, and Lihua Xie. 2021. Range-focused fusion of camera-IMU-UWB for accurate and drift-reduced localization. *IEEE Robot. Automat. Lett.* 6, 2 (Apr. 2021), 1678–1685.

[29] Thien-Minh Nguyen, Thien Hoang Nguyen, Muqing Cao, Zhirong Qiu, and Lihua Xie. 2019. Integrated UWB-vision approach for autonomous docking of UAVs in GPS-denied environments. In *Proceedings of the International Conference on Robotics and Automation (ICRA'19)*. 9603–9609.
[30] Edwin Olson. 2011. AprilTag: A robust and flexible visual fiducial system. In *Proceedings of the International Conference on Robotics and Automation (ICRA'11)*. 3400–3407.
[31] Francisco Javier Perez-Grau, Fernando Caballero, Luis Merino, and Antidio Viguria. 2017. Multi-modal mapping and localization of unmanned aerial robots based on ultra-wideband and RGB-D sensing. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots Systems (IROS'17)*. 3495–3502.
[32] Tong Qin, Shaozu Cao, Jie Pan, and Shaojie Shen. 2019. A general optimization-based framework for global pose estimation with multiple sensors. *arXiv:1901.03642* (2019).
[33] Tong Qin, Peiliang Li, and Shaojie Shen. 2018. VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Trans. Robot.* 34, 4 (Aug. 2018), 1004–1020.
[34] Jorge Peña Queralta, Carmen Martínez Almansa, Fabrizio Schiano, Dario Floreano, and Tomi Westerlund. 2020. UWB-based system for UAV localization in GNSS-denied environments: Characterization and dataset. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots Systems (IROS'20)*. 4521–4528.
[35] Priya Roy and Chandreyee Chowdhury. 2022. A survey on ubiquitous WiFi-based indoor localization system for smartphone users from implementation perspectives. *CCF Trans. Pervasive Comput. Interact.* 4, 3 (Jan. 2022), 298–318.
[36] Dinesh Kumar Sah, Tu N. Nguyen, Manjusha Kandulna, Korhan Cengiz, and Tarachand Amgoth. 2022. 3D localization and error minimization in underwater sensor networks. *ACM Trans. Sensor Netw.* 18, 3 (2022), 1–25.
[37] Cung Lian Sang, Michael Adams, Marc Hesse, and Ulrich Rückert. 2023. Bidirectional UWB localization: A review on an elastic positioning scheme for GNSS-deprived zones. *IEEE J. Indoor Seaml. Posit. Navig.* 1 (2023), 161–179.
[38] Yang Song and Li-Ta Hsu. 2021. Tightly coupled integrated navigation system via factor graph for UAV indoor localization. *Aerosp. Sci. Technol.* 108, 106370 (Jan. 2021).
[39] Pieter van Goor and Robert Mahony. 2023. EqVIO: An equivariant filter for visual-inertial odometry. *IEEE Trans. Robot.* (2023). doi:10.1109/TRO.2023.3289587.
[40] Chen Wang, Handuo Zhang, Thien-Minh Nguyen, and Lihua Xie. 2017. Ultra-wideband aided fast localization and mapping system. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots Systems (IROS'17)*. 1602–1609.
[41] Jun Wang, Pengfei Gu, Lei Wang, and Ziyang Meng. 2024. RVIO: An effective localization algorithm for range-aided visual-inertial odometry system. *IEEE Trans. Intell. Transp. Syst.* 25, 2 (Feb. 2024), 1476–1490.
[42] Yingying Wang, Hu Cheng, and Max Q.-H. Meng. 2022. A2DIO: Attention-driven deep inertial odometry for pedestrian localization based on 6D IMU. In *Proceedings of the International Conference on Robotics and Automation (ICRA'22)*. 819–825.
[43] Junyu Wei, Haowen Wang, Shaojing Su, Ying Tang, Xiaojun Guo, and Xiaoyong Sun. 2022. NLOS identification using parallel deep learning model and time-frequency information in UWB-based positioning system. *Measurement* 195 (May 2022), 111191.
[44] Jingao Xu, Erqun Dong, Qiang Ma, Chenshu Wu, and Zheng Yang. 2021. Smartphone-based indoor visual navigation with leader-follower mode. *ACM Trans. Sensor Netw.* 17, 2 (2021), 1–22.
[45] Bo Yang, Jun Li, and Hong Zhang. 2021. Resilient indoor localization system based on UWB and visual-inertial sensors for complex environments. *IEEE Trans. Instrum. Meas.* 70 (2021), 1–14.
[46] Junjie Yin, Zheng Yang, Sicong Liao, Chunhui Duan, Xuan Ding, and Li Zhang. 2023. TagFocus: Towards fine-grained multi-object identification in RFID-based systems with visual aids. *ACM Trans. Sensor Netw.* 19, 1 (2023), 1–22.
[47] Di Yuan, Xiaojun Chang, Zhihui Li, and Zhenyu He. 2022. Learning adaptive spatial-temporal context-aware correlation filters for UAV tracking. *ACM Trans. Multimedia Comput. Commun. Appl.* 18, 3 (2022), 1–18.
[48] Faheem Zafari, Athanasios Gkelias, and Kin K. Leung. 2019. A survey of indoor localization systems and technologies. *IEEE Commun. Surveys Tuts.* 21, 3 (3rd Quarter. 2019), 2568–2599.
[49] Shibin Zhao, Ji Li, and Yanbo Zhang. 2023. Loosely coupled localization system based on multi-base station UWB and point-line VIO. In *Proceedings of the International Conference on Electronic Information Engineering and Computer Science (EIECS'23)*. IEEE, 136–142.
[50] S. Zheng, Z. Li, Y. Liu, H. Zhang, P. Zheng, X. Liang, Y. Li, X. Bu, and X. Zou. 2022. UWB-VIO fusion for accurate and robust relative localization of round robotic teams. *IEEE Robot. Automat. Lett.* 7, 4 (Oct. 2022), 11950–11957.

Received 27 August 2023; revised 28 March 2024; accepted 29 April 2024