DTVIR-Swarm: A Distributed and Tightly Integrated Visual-Inertial-UWB System for Anchor-Free Swarm Cooperative Localization

Authors: Xincan Luo

Date: 2025-05-15T21:40:13+00:00

Abstract

Accurate UAV positioning is vital for swarm cooperation. However, this remains challenging in situations where GNSS and other external infrastructures are unavailable. To address the challenge, we propose to only use the onboard MIMU, monocular camera and UWB device to construct a distributed and anchor-free cooperative localization system by tightly fusing the measurements. As the onboard UWB measurements in dynamic motion conditions is noisy and discontinuous, we propose an adaptive adjustment method based on chi-square detection to effectively filter out inconsistent and false ranging information. Moreover, we introduce the pose-only theory to model the visual measurement, which improves the efficiency and accuracy for visual-inertial processing. Then a sliding window extend Kalman filter was constructed to fuse all the measurements in a tightly way, which is capable to work under UWB or visual deprived conditions. To overcome the state consistency challenge inherent in the distributed cooperative structure, we propose to not only model the UWB noisy uncertainty but also the neighbor agent's position uncertainty into the measurement model. To validate the effectiveness of proposed our methods, we have established both simulation and hardware test platforms. The proposed method is compared with state-of-the-art (SOTA) UAV localization approaches designed for GNSS-challenged environments. Extensive experiments demonstrate that our algorithm achieves superior positioning accuracy and higher computational efficiency. Moreover, even when vision loss causes other methods to fail, our proposed method continues to operate effectively.

Full Text

Preamble

Available online at www.sciencedirect.com
Procedia Economics and Finance 00 (2012) 000–000

DTVIR-Swarm: A Distributed and Tightly Integrated Visual-Inertial-UWB System for Anchor-Free Swarm Cooperative Localization

Xincan Luo[12], Xueyu Du[12], Shuai Yue[12], Yunxiao Lv[12], Lilian Zhang[12], Xiaofeng He[12], Wenqi Wu[12], and Jun Mao[12*]

[1]College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, Hunan, China
[2]National Key Laboratory of Equipment State Sensing and Smart Support, Changsha 410073, Hunan, China

Abstract

Accurate UAV positioning is vital for swarm cooperation, yet remains challenging in situations where GNSS and other external infrastructures are unavailable. To address this challenge, we propose a distributed, anchor-free cooperative localization system that uses only onboard MIMU, monocular camera, and UWB devices, tightly fusing their measurements. Since onboard UWB measurements under dynamic motion conditions are noisy and discontinuous, we introduce an adaptive adjustment method based on chi-square detection to effectively filter out inconsistent and false ranging information. Moreover, we employ pose-only theory to model visual measurements, which improves the efficiency and accuracy of visual-inertial processing. A sliding window extended Kalman filter is constructed to tightly fuse all measurements, enabling operation under vision-deprived or UWB-deprived conditions. To overcome the state consistency challenge inherent in distributed cooperative structures, we model both UWB measurement noise uncertainty and neighbor agent position uncertainty into the measurement model. We validate our methods through both simulation and hardware test platforms, comparing the proposed system against state-of-the-art UAV localization approaches designed for GNSS-challenged environments. Extensive experiments demonstrate that our algorithm achieves superior positioning accuracy and higher computational efficiency. Furthermore, even when vision loss causes other methods to fail, our proposed method continues to operate effectively.

Keywords: Cooperative Localization, Multi-UAV, UWB Ranging, Sensor Fusion

*Corresponding author. E-mail address: maojun12@nudt.edu.cn.

1. Introduction

Accurate localization is critical for efficient operations and precise control in aerial robotics, particularly in multi-UAV systems where collaborative tasks, formation flying, and mission execution depend on reliable positioning. However, achieving reliable localization remains challenging in GNSS-denied environments. To overcome this limitation, researchers have investigated multimodal sensor fusion approaches, among which visual-inertial odometry (VIO) has emerged as a widely studied method that integrates camera and IMU data for relative motion estimation and localization [1-3].

VIO operates by capturing environmental images through cameras to extract and track visual features, while IMUs measure acceleration and angular velocity to estimate pose changes. Despite its advantages, VIO suffers from two primary limitations. First, its performance heavily depends on visual conditions: feature extraction becomes unreliable in texture-sparse areas, and sudden environmental changes introduce tracking errors that degrade localization accuracy. Second, the computational complexity of real-time image processing and sensor fusion strains the limited processing capabilities of UAVs, further restricting accuracy and practicality [4-6].

To address these challenges in complex environments, multi-UAV collaborative localization methods have been developed. These approaches are broadly classified into centralized and distributed systems based on data processing strategies. Centralized methods rely on transmitting all UAV-collected data to a ground station or central node for unified computation. However, they require stable network connectivity and high-performance hardware at the central node, resulting in scalability and flexibility limitations. In contrast, distributed systems eliminate central nodes by enabling direct communication between neighboring UAVs, offering enhanced robustness and adaptability for dynamic environments. This advantage has made distributed methods increasingly prevalent in multi-UAV applications [7,8].

Current VIO-based collaborative localization methods often improve accuracy by sharing environmental features observed across UAV clusters. While effective, such strategies impose high communication bandwidth requirements, which become problematic as UAV numbers or environmental complexity increase. Alternative approaches using pre-deployed UWB anchors enhance localization through anchor-based distance measurements. Nevertheless, these methods lack flexibility due to their dependence on carefully calibrated anchor infrastructure, making them impractical for unknown or dynamic environments where anchor deployment is costly or infeasible [9,10].

To address these challenges and meet the demands for low cost, lightweight design, and real-time performance, this paper proposes a distributed anchor-free visual-inertial-UWB collaborative localization system for multi-UAVs based on the sliding window extended Kalman filter framework. In this system, UAVs measure mutual distances using UWB technology without relying on UWB anchors. To fully utilize UWB measurements, multiple UWB keyframes are incorporated into the sliding window, and an adaptive adjustment method for UWB filtering errors based on chi-square detection is employed to ensure robustness and real-time performance. Additionally, to further enhance accuracy, a visual observation model based on pose-only (PO) theory is adopted, improving the system's adaptability to challenging visual environments [11,12].

Specifically, the contributions of this paper are as follows:

- We have developed a distributed and anchor-free cooperative localization system for multi-UAVs. In this system, each UAV leverages its onboard sensors to estimate its own state in a distributed manner, eliminating the need for a central node or anchor node.
- We propose novel processing methods for both UWB and visual measurements. To address noisy onboard UWB measurements, we integrate multiple UWB keyframes into the sliding window and propose an adaptive error adjustment method using chi-square detection to boost the algorithm's robustness and accuracy. Moreover, we apply pure pose (PO) theory to model multi-view visual measurements, enhancing the algorithm's accuracy and efficiency.
- We have developed a multi-sensor fusion framework using a sliding window EKF to tightly combine inertial, visual, and UWB ranging measurements. Thanks to this new tightly integrated localization structure, the system can effectively handle situations with vision loss, providing a more robust solution for cooperative localization in environments where vision may be temporarily unavailable.
- We construct a multi-UAV system for extensive experiments. A set of miniaturized collaborative localization system hardware was built and adapted to the algorithm proposed in this paper. Experiments conducted in real-world scenarios proved the lightweight design of the system and its applicability to practical applications.

The structure of the paper is as follows. The second part is a literature review. The third part introduces the proposed algorithm and the detailed mathematical model of collaborative localization. The fourth part presents simulation and real-world experiments with results analysis. Finally, the fifth part provides the conclusion.

2. Related Work

In the field of UAV collaborative localization, a key issue is the relative position observation between UAVs. Currently, there are many methods to solve this problem of relative position observation between drones. Marker-based visual mutual observation methods usually extract and deploy markers on the drone, using markers such as ultraviolet LED lights and ultraviolet-sensitive cameras combined to perform relative positioning [13]. Unlabeled visual mutual observation methods often rely on convolutional neural networks (CNN) [14-16], which use machine learning-based techniques to extract relative distances, but are easily affected by changes in the appearance of targets and the environment and cannot provide accurate relative estimates. Multi-robot SLAM methods can use map merging and loop closure between robots to obtain relative attitudes, but these methods require the communication of large amounts of data and are not suitable for drone platforms with computational performance constraints.

3D LiDAR and UWB can also directly measure relative distances. In [18], the fusion of 3D LiDAR, fisheye camera, and UWB data is used to track drones flying above LiDAR-equipped unmanned ground vehicles (UGVs), where drones wirelessly share IMU and 3D LiDAR data. In [19], a distributed LiDAR inertial group range measurement method was proposed, which uses reflectivity values from LiDAR data to directly detect collaborative drones, but using LiDAR as a relative position observation sensor is costly. In [20], a relative positioning method for micro-drones based on VIO and LiDAR localization was proposed. Slave drones use LiDAR to observe the relative position and distance of the master drone and integrate it with VIO to improve the positioning accuracy of slave drones. However, this method requires a high-precision master drone node; if the master drone node is damaged, the entire drone cluster will not work normally.

To address substantial positioning errors in GNSS-challenged environments, particularly when visual features are sparse, researchers have proposed several methods. In terms of single-drone positioning, document [21] fuses data from static UWB anchors, LiDAR odometry, IMUs, and VIOs to improve drone positioning accuracy in GNSS-challenged environments, but in emergency situations, it is not feasible to place static UWB anchors in the area. In [22], a multi-sensor framework is proposed that uses ultra-wideband (UWB) technology and visual-inertial measurement (VIO) to provide robust and low-drift positioning. In [23], a learning-based drone positioning method is proposed that uses fused vision, IMU, and UWB sensors, combined with visual-inertial (VIO) and UWB branches, to predict global attitude, but these two methods still require the placement of multiple ground UWB anchors in advance.

In terms of multi-drone positioning, document [24] proposed a method that integrates UWB and VIO for collaborative positioning of two drones, and document [25] proposed a method for distributed formation estimation in large drone clusters. In [8], a distributed collaborative SLAM system was proposed, which innovatively manages near-field estimation and integrates multiple sensors and map data to achieve accurate near-field relative state estimation and consistent global trajectory far-field estimation. In [26], the authors fused detection results from CNN with UWB data and VIO for relative positioning in drone clusters. Document [27] fuses data from static UWB anchors, mutual ranging between drones, IMU, and VIO to improve drone positioning accuracy in GNSS-challenged environments.

Zhang [28] fused active vision-based relative positioning with UWB and VIO data for formation control of tagged drone clusters. However, this active vision-based approach cannot cope with urgent tasks. The authors of document [29] focused on collaborative localization of UGV and UAV teams, and their method relied mainly on UWB and VIO data and used 3D LiDAR detection during initialization. The authors of document [30] utilized a heterogeneous team of UGVs carrying LiDAR and camera-equipped drones with the goal of detecting UGVs from airborne cameras on the drones and using them as landmarks for improving drone positioning.

Document [31] proposed a new cooperative localization framework based on optimized belief propagation (BP) for use in GNSS-denied areas, and document [32] proposed a technology to improve the accuracy of visual-inertial odometry (VIO) by combining ultra-wideband (UWB) positioning technology. However, these methods all require the assistance of multiple UWB anchors, which is not conducive to practical use. Document [33] proposes a multi-drone relative scheme based on distributed graph optimization (DGO), which combines onboard ultra-wideband (UWB) modules, cameras, and inertial sensors, but the cameras need to observe other drones for position estimation, making it unsuitable for large-scale scenarios. Document [34] proposes a collaborative positioning algorithm for unmanned aerial vehicles based on improved sequential quadratic programming (SQP) and unscented Kalman filter (UKF). However, in actual use, IMU sensors need high accuracy and high cost.

Therefore, considering the requirements of low cost, lightweight design, and real-time performance, we propose a distributed anchorless vision-inertial-UWB multi-drone collaborative positioning system. UAVs measure the distance between drones through UWB without relying on UWB anchors, and a filtering framework based on EKF is used to fuse multiple observations, which enhances the lightweight and real-time nature of the algorithm. To further improve accuracy, we incorporate multiple UWB keyframes into the sliding window and adopt an adaptive adjustment method for filtering errors based on chi-square detection. In addition, in terms of vision observation, we adopt a visual observation model based on PO constraints, which improves the system's adaptability to challenging visual environments.

3. Method

The proposed method is a distributed and anchor-free visual-inertial-UWB cooperative localization system for UAVs. It consists of three main components: a sliding window extended Kalman filter framework for tightly fusing inertial, visual, and UWB ranging data; an adaptive adjustment method for UWB measurements based on chi-square detection; and a visual observation model using pose-only (PO) theory. The following sections detail these components.

3.1. Overview of the Architecture

The block diagram of the architecture of the proposed system is shown in Fig. 1 [Figure 1: see original paper]. Each UAV is equipped with an IMU, a camera, and a UWB sensor, and the UAVs can share each other's position and relative distance information through communication links and UWB.

In the structural block diagram shown in Figure 1, feature tracking and extraction are used to obtain environmental feature information, IMU data is used for state propagation and enhancement to predict the motion state of the UAV, and basic views and sensor data are used to build a visual observation model based on pure pose (PO) theory to improve the accuracy of the algorithm. Then, the information predicted by the filter is used to perform chi-square detection on the UWB data and filter out outliers. Finally, the state estimate of the UAV is updated by fusing multiple sensor information in a tight combination method.

3.2. Filter State and Propagation

In the proposed multi-UAV collaborative navigation system, we employ the sliding window EKF for state estimation, which is developed from OpenVINS [36]. The state of any drone in the drone cluster is defined as follows: The error state vector δx consists of the current INS error parameter Iδx, vision measurement keyframes, and UWB measurement keyframes. When the lengths of the visual sliding window and the UWB sliding window are N and M respectively, they can be defined as:

δx = [Iδx, δx_{visual}, δx_{uwb}]

where the current inertial localization error state Iδx includes the attitude error δ , position error δp, velocity error δv, gyroscope bias error δb_g, and accelerometer bias error δb_a. Vision measurement keyframe error and UWB measurement keyframe error are defined as follows:

δx_{visual} = [δ _b$_1$, δp_b$_1$, ..., δ _{b_N}, δp_{b_N}] δx_{uwb} = [δ _b$_1$, δp_b$_1$, ..., δ _{b_M}, δp_{b_M}]

where δ _b and δp_b are the IMU attitude and position errors at the n-th vision keyframe time; δ _b and δp_b are the IMU attitude and position errors at the m-th UWB keyframe time. The true state x can be obtained from the estimated state $\hat{x}$ and the error state δx:

x = $\hat{x}$ δx

For the attitude error, the operator   is given by the quaternion composition. For other states, the operator   is equivalent to Euclidean addition. When IMU measurements are available, INS mechanization is conducted to output high-frequency prior poses. Meanwhile, the forward propagation of the whole error state and its covariance is similar to OpenVINS [36] and will not be repeated here. When a new visual/UWB keyframe is added to the sliding window, state augmentation and covariance update are needed to incorporate the visual/UWB state into the state vector, and the corresponding covariance P is augmented as:

$$P \leftarrow J\ [P\ 0;\ 0\ P_{new}]\ J$$

where J is the Jacobian matrix representing the relationship between the newly added state and the original state. When the visual/UWB sliding window exceeds its maximum length, marginalization occurs and the state and covariance of the oldest visual/UWB keyframe are directly deleted [36].

3.3. Visual Measurement Based on PO Theory

To simplify the complexity of the 3D reconstruction process in traditional VIO systems, reduce computational load, and avoid accuracy limitations imposed by direct 3D reconstruction, we reconstruct the measurement model using only pixel coordinates and relative positional pose based on PO theory [37-39]. In PO theory, multi-view geometry can be described using only camera poses, meaning that instead of directly estimating the 3D coordinates of feature points in the scene, we infer their positional relationships from the relative poses between cameras.

Assuming that the projection of feature point p in image i can be represented as $z_i = \pi(K[R_i|t_i]P)$, where K is the camera intrinsic matrix, R_{ij} and t_{ij} are the rotation and translation from camera i to camera j, respectively, the reprojection error using the 3D coordinates of feature points can be represented in PO theory without directly reconstructing 3D points. The geometric description of multiple views based on PO theory can be expressed as:

$$d_{jk} = d_{ji} + d_{ik}$$

where d_{ji} represents the constraint between images i and j, Cp_i are the normalized coordinates of feature points in image i, and R_{ij} and t_{ij} are the rotation and translation from image i to image j, respectively. Thus the reprojection error can be redefined as:

$$e_l^i = z_l^i - \pi(K[R_i|t_i]Cp_j)$$

where e_l^i is the reprojection error of the l-th feature point in the i-th image, and z_l^i is the projection obtained by using only the camera pose and 2D features. The measurement model based on PO theory can be represented as follows [37-39]:

$$z = h(x) + n$$

where Te is a transformation vector for converting 3D vectors to 2D, H is the measurement Jacobian matrix, and x_n is the state vector. As a result, the new measurement model is represented only in pixel coordinates and system attitude, fully decoupled from 3D features, avoiding the effects of inaccurate 3D reconstruction processes.

3.4. Collaborative Localization with Anchor-Free UWB Measurement

The scheme of UAVs using UWB measurements for cooperative localization is shown in Fig. 2 [Figure 2: see original paper]. Ranging observations between UAVs via UWB can be used to obtain information about the distances to and from other UAVs within communication range and the positions of other UAVs, which are utilized to construct quantitative equations. Taking the example of mutual ranging between UAV i and UAV j, the distance measurement for UAV i can be written as:

đ*{ij} = ‖Gp_i - Gp_j‖ +* d{ij}

where Gp_i and Gp_j represent the position vectors of UAV i and UAV j in the global coordinate system, respectively, đ*{ij} represents the distance of UAV i relative to UAV j measured by UWB, and* d{ij} is the distance measurement difference. The distance measurement matrix of UAV i relative to UAV j can be defined by the following partial derivatives:

H_{ij} = đ_{ij}/ x = [đ*{ij}/ Gp_i,* đ{ij}/ Gp_j]

Since UAV j does not have accurate a priori position results, its position is also obtained from the distributed estimator, so it will impact UWB range measurements, and its uncertainty needs to be modeled in the distributed collaborative localization system, as shown in the following formula:

Σ_{ij} = H_{ij} P H_{ij} + R_d + J_j P_j J_j

where Σ_{ij} is the innovation covariance, P is the error covariance of UAV i, R_d is the UWB measurement noise, J_j is the Jacobian with respect to UAV j's position, and P_j is the position error covariance of UAV j. Therefore, the variance of UWB measurement d_{ij} caused by neighbor drone node positioning errors is:

σ^2_{ij} = J_j P_j J_j

Then the observation noise matrix of UWB ranging measurement is:

R = R_d + σ^2_{ij}

where δd is the ranging error of the UWB sensor, and GpP is the position error covariance of UAV j. We can obtain the measurement equation for cooperative navigation based on UWB measurements:

z = h(x) + v

To improve the accuracy and reliability of the system, outlier errors of UWB measurements in the Kalman filtering system are filtered out using the chi-square test. The pseudo-code is shown in Algorithm 1. Since the residual vector obeys a zero-mean Gaussian distribution, the normalized sum of squares of its values should obey a chi-square distribution with degrees of freedom equal to the dimensions of the residual vector. However, outliers in UWB measurements undermine this assumption, so we can define the chi-square distribution statistics and criteria as follows:

Algorithm 1: UWB Range Update with Chi-Square Test

Input: Prior state estimate x̂_k$^-$ and covariance P_k$^-$
Output: Posterior state estimate x̂_k and covariance P_k

1. Predict measurement: ż = h(x̂_k$^-$)

2. Compute innovation (residual): y = z - ż

3. Compute innovation covariance matrix and chi-square statistic: S = H P_k$^-$ H + R, ^{2} = y S^{-1} y

4. Compare chi-square statistic with threshold: if ^{2} > ^{2}_{threshold}, reject measurement; else proceed with update

where P_k$^-$ is the a priori error covariance matrix of the Kalman filter, and R is the observation noise covariance matrix. In this article, the threshold of the chi-square test is set to 95%, and the dimension of the residual vector is 6, so the corresponding threshold is 12.59. Values above this threshold will be marked as outliers and will not be added to the UWB measurement sliding window. Finally, each UAV uses the standard EKF update formula to update UWB measurements and obtain state estimates after collaborative positioning.

3.5. Measurement Tight Fusion

The framework for tight fusion of multi-sensor information proposed in this paper is shown in Fig. 3 Figure 3: see original paper, while other multi-sensor information fusion frameworks based on optimized cluster positioning methods are shown in Fig. 3(a).

As depicted in Fig. 3, our approach in Fig. 3(b) demonstrates greater autonomy in information acquisition and superior fusion efficiency. We independently procure data from the IMU, VIO, and UWB systems. The IMU operates autonomously, without reliance on external environmental factors. Consequently, even if vision or UWB measurements encounter short-term issues, the system can continue functioning by leveraging IMU data. After undergoing processing by dedicated modules, information from each sensor is integrated within the measurement tight fusion module. This integration combines the strengths of each sensor, thereby enhancing positioning accuracy and system stability. In contrast, other methods, such as those illustrated in Fig. 3(a) (e.g., Literature 1 and Literature 2), employ joint optimization and depend on the VIO system to function properly. These systems face complete failure if the VIO system malfunctions. By tightly fusing measurements, our method substantially enhances system robustness and adaptability, providing a significant advantage in complex environments.

4. Experimental Evaluation

This section provides a comprehensive performance evaluation of our proposed collaborative navigation framework and benchmarks it against state-of-the-art VIO methods and collaborative positioning methods commonly used for drone positioning in satellite-denied situations. Experimental validation was conducted using datasets collected from both the AirSim simulation platform and real-world environments. To quantitatively assess localization accuracy, we employed the EVO toolkit to compute absolute pose error (APE) metrics along complete trajectories. All evaluations were performed on a laptop configuration (13th Gen Intel® Core™ i9-13900HX 2.20 GHz processor) to ensure practical applicability.

4.1. Simulation Experiment

To verify the effectiveness of the new architecture, we first conducted simulation experiments based on the AirSim platform. Since the AirSim platform does not have UWB data simulation, we added Gaussian white noise with a standard deviation of 0.3m based on the true distance to simulate UWB data. The parameter settings of the simulation experiment are shown in Table 1 , and the simulation scenarios and ground-truth trajectories are shown in Fig. 4 [Figure 4: see original paper].

Table 1. Sensors carried by the UAV platform in simulation experiments.

Sensors	Frequency	Parameters
Camera (UAV1&UAV2&UAV3)	100Hz	Resolution: 752$×$480 pixels, View angle: 90°
IMU	200Hz	In-run bias stability of Gyroscope: 1°/h, Noise density of Gyroscope: 0.1°/s/√Hz, In-run bias stability of Accelerometer: 2000 µg, Noise density of Accelerometer: 200 µg/√Hz
UWB	10Hz	Range accuracy: 0.3m, Maximum range: 500m

Fig. 4. (a) UAV flight environment in simulation experiments. (b) Motion trajectories of the three UAVs in simulation experiments.

In the simulation experiment, the positioning trajectory of the drone is shown in Fig. 5 [Figure 5: see original paper], where "gt" represents the ground-truth trajectory, "vins_{mono}" and "open_{vins}" represent two state-of-the-art VIO methods from literature [35] and [36], "coo_{vir}" represents the collaborative localization method proposed in literature [27], and "coo_{our}" represents the method proposed in this paper. The statistics of drone positioning root-mean-square error (RMSE) results are shown in Table 2 . Since this paper targets drones operating in large-scale outdoor environments requiring real-time positioning information output, loop closure repositioning is not considered when comparing with OpenVINS and VINS-Mono algorithms. Additionally, since the OpenVINS, VINS-Mono, and VIR_{SLAM} algorithms do not have initial heading alignment functionality, we rotate the trajectories so that the initial heading is consistent with the true value.

Fig. 5. Comparison of estimated trajectories of three algorithms with ground truth in simulation experiments, where "coo" represents the method proposed in this paper.

Table 2. RMSE of the position in simulation experiments, in meters.

Method	UAV1	UAV2	UAV3	Average
VINS-MONO	3.42	3.51	3.38	3.44
Open-VINS	3.41	3.52	3.39	3.44
COO_{VIR}	2.85	2.92	2.88	2.88
COO(ours)	2.07	2.13	2.09	2.10

It can be seen from Table 2 that in the simulation experiment, compared with the VINS-Mono algorithm, the average positioning error of the drone is reduced by 39.4%; compared with the OpenVINS algorithm, the average positioning error is reduced by 39.5%; and compared with the VIR_{SLAM} collaborative positioning algorithm, the average positioning error is reduced by 25.4%. Therefore, simulation results show that better positioning performance can be achieved using our proposed collaborative localization algorithm.

4.2. Vision Loss Experiment

To verify the efficacy of the proposed algorithm in scenarios of visual loss, we conducted experiments by introducing partial visual loss into the simulation framework established in the previous section, as illustrated in Figure 6 [Figure 6: see original paper], with each episode of visual loss lasting for a duration of 10 seconds. Unless otherwise specified, the experimental settings, including parameters, configurations, and environmental conditions, were consistent with those outlined in the preceding simulation experiment.

Fig. 6. Trajectory containing visual loss periods for UAV1, UAV2, and UAV3.

Fig. 7. Comparison of estimated trajectories with ground truth in vision loss simulation experiments, where "coo" represents the method proposed in this paper.

In the simulated vision loss experiment, the UAV's positioning trajectory is shown in Fig. 7, where "gt" denotes the ground-truth trajectory and "coo_{vir}" represents the collaborative localization method based on joint optimization from literature [27]. It can be seen that the collaborative localization method based on joint optimization relies on a continuous VIO system. When vision is lost, the VIO system fails to function properly, causing significant deviation in the overall system's positioning results and rendering it unusable. In contrast, our proposed method can effectively handle vision deprivation, maintaining accuracy comparable to that before vision loss.

4.3. Physical System Construction

To verify the practical application capabilities of the algorithm, this research built a distributed multi-drone collaborative localization hardware platform. The collaborative localization platform integrates multi-modal sensors, carries out clock synchronization, and includes a built-in embedded computing module that can process sensor data in real time. The platform can be mounted on a small quad-rotor drone and realizes status information sharing and UWB ranging data interaction through a wireless communication module.

Fig. 8. Collaborative localization physical system hardware platform.

The hardware architecture is shown in Fig. 8, which mainly includes the following modules: (1) Vision module: Equipped with FLIR Blackfly S USB3 global shutter camera, achieving microsecond time synchronization between IMU and camera through hardware trigger signals. (2) Inertial module: Adopts MEMS-IMU to output raw data at 200Hz. (3) UWB module: The ranging frequency is 10Hz, achieving accuracy of 0.3m in dynamic environments through TDoA technology (LOS conditions). (4) RTK positioning module: Integrated RTK-GNSS module (positioning accuracy of 2cm) serves as ground truth and is only enabled during algorithm evaluation. (5) Main processor: Selects the RK3588 embedded computing module to meet the real-time computing requirements of the sliding window EKF framework.

The parameters of each module are shown in Table 3 .

Table 3. Sensors carried by the UAV platform in real experiments.

<table>
<tr><th>Sensors</th><th>Frequency</th><th>Parameters</th></tr>
<tr><td>Camera (UAV1&UAV2)</td><td>30Hz</td><td>Resolution: 752$×480pixels, LensFocalLength : 8mm||Camera(UAV3)|30Hz|Resolution : 1280×$1024 pixels, Lens Focal Length: 6mm</td></tr>
</table>

Sensors	Frequency	Parameters
IMU	200Hz	In-run bias stability of Gyroscope: 10°/h, Noise density of Gyroscope: 0.8°/s/√Hz, In-run bias stability of Accelerometer: 5000 µg, Noise density of Accelerometer: 500 µg/√Hz
UWB	10Hz	Range accuracy: <0.5m, Maximum range: 500m
GNSS(RTK)	10Hz	Positioning accuracy: <0.1m

4.4. Real-World Experiment

To prove the practicality of the system proposed in this paper in real-world scenarios, we conducted a real-world collaborative localization experiment with three drones. The flight trajectory of the drones is shown in Fig. 9 Figure 9: see original paper. The sensor parameters carried by the drone are shown in Table 3. The experimental environment is an outdoor scene, as shown in Fig. 9(a). In the real-world experiment, the images captured by the drone camera are shown in Fig. 10 [Figure 10: see original paper]. It can be seen that compared with drones flying indoors, when the drone is performing outdoor missions with the camera facing down, it sees fewer feature points in the picture. The drones are equipped with high-precision RTK positioning equipment to provide ground truth position values.

Fig. 9. (a) UAVs in the experiment. (b) Motion trajectories of the three UAVs (RTK). (c) Comparison of UWB measured distance and real distance between UAV 1 and UAV 2.

Fig. 10. Images captured by UAV cameras in real-world experiments.

To verify several innovative aspects of the method proposed in this paper, particularly the novel processing methods for both UWB and visual measurements, ablation experiments with different settings were also carried out.

The positioning errors of the three drones are compared and analyzed using high-precision RTK as ground truth. The results are shown in Fig. 11 [Figure 11: see original paper].

Fig. 11. Comparison of estimated trajectories of three algorithms with ground truth in real-world experiments, where "Ours" represents the method proposed in this paper.

Fig. 12. Runtime comparison of different algorithms in real-world experiments, in milliseconds.

Table 4. RMSE of the position in real-world experiments, in meters.

Method	UAV1	UAV2	UAV3	Average
VINS-MONO	2.85	2.92	2.78	2.85
OpenVINS	2.83	2.90	2.76	2.83
COO_{{{VIR}}_{{SLAM}}}	2.12	2.18	2.05	2.12
COO(ours_{{{without}}_{{PO}}})	1.85	1.92	1.78	1.85
COO(ours_{{{without}}*{{UWB}}}*{SW})	1.92	1.98	1.85	1.92
COO(ours_{{{without}}*{{UWB}}}*{CT})	1.88	1.95	1.82	1.88
COO(ours)	1.42	1.48	1.35	1.42

Table 4 shows a detailed evaluation of positioning errors for real-world experiments, where “COO_{{{VIR}}*{{SLAM}}}” represents the collaborative localization method proposed in document [27], ”COO(ours)” represents the algorithm proposed in this paper, ”COO(ours*{{{without}}*{{PO}}})” represents the algorithm proposed in this paper without using PO constraints, ”COO(ours*{{{without}}*{{UWB}}}*{SW})” represents the algorithm without using UWB sliding windows, and “COO(ours_{{{without}}*{{UWB}}}*{CT})” represents the algorithm without using UWB chi-square detection. As we propose novel processing methods for both UWB and visual measurements, our algorithm can achieve better positioning accuracy in the outdoor scenes in this paper. Fig. 12 shows a comparison of the running times of different algorithms to evaluate operating efficiency. It can be seen that our algorithm can maintain high operating efficiency thanks to its lightweight architecture and filter-based advantages. To sum up, compared with open-source state-of-the-art VIO positioning methods and multi-UAV collaborative positioning methods, the algorithm proposed in this paper shows excellent positioning performance.

5. Conclusion and Discussion

This paper addresses the challenge of degraded navigation and positioning accuracy in GNSS-challenged environments for unmanned platforms, where visual environmental features are sparse and GNSS communication is constrained. We propose a distributed anchor-free visual-inertial-UWB multi-UAV cooperative localization system. In this system, UWB measurements between UAVs are utilized to estimate inter-drone distances without relying on pre-deployed UWB anchors. A sliding window EKF framework is adopted to fuse multimodal observations, enhancing the algorithm’s lightweight design and real-time performance. To further improve accuracy, a PO visual observation model is introduced, which strengthens the system’s adaptability to challenging visual environments with limited features. Additionally, multiple UWB keyframes are integrated into the sliding window, and a chi-square test-based adaptive adjustment method is employed to dynamically optimize filtering errors, ensuring robustness and precision. Different from current vision-inertial-UWB cooperative localization systems based on the VIO system, we adopt a tightly-coupled cooperative localization system based on IMU, which can effectively deal with vision loss. Simulation and real-world experiments demonstrate that the proposed method outperforms state-of-the-art localization approaches in GNSS-denied scenarios, achieving higher positioning accuracy. Moreover, our method can effectively handle situations of vision loss. Furthermore, the system has been successfully deployed on small-scale UAV platforms, validating its practicality and applicability in real-world lightweight drone systems.

In future research, we will further explore the application of our system in challenging environments such as large-scale scenarios and day-night transitions to obtain a more comprehensive and powerful drone cluster positioning system.

References

[1] Al Mahmud, S., Kamarulariffin, A., Ibrahim, A.M., Mohideen, A.J.H., 2024. Advancements and Challenges in Mobile Robot Navigation: A Comprehensive Review of Algorithms and Potential for Self-Learning Approaches. J. Intell. Robot. Syst. https://doi.org/10.1007/s10846-024-02149-5.

[2] Cai, D., Li, S., Qi, W., Ding, K., Lu, J., Liu, G., Hu, Z., 2024. DFT-VSLAM: A Dynamic Optical Flow Tracking VSLAM Method. J. Intell. Robot. Syst. https://doi.org/10.1007/s10846-024-02171-7.

[3] Si, J., Li, B., Wang, L., Deng, C., Wang, J., Wang, S., 2024. A UAV Autonomous Landing System Integrating Locating, Tracking, and Landing in the Wild Environment. J. Intell. Robot. Syst. https://doi.org/10.1007/s10846-023-02041-8.

[4] Fu, G., Wang, Y., Yang, J., Wang, S., Yang, G., 2023. Monocular Visual Navigation Algorithm for Nursing Robots via Deep Learning Oriented to Dynamic Object Goal. J. Intell. Robot. Syst. https://doi.org/10.1007/s10846-023-02024-9.

[5] Zhou J, Gu G, Chen X. Distributed Kalman filtering over wireless sensor networks in the presence of data packet drops[J]. IEEE Transactions on Automatic Control, 2018, 64(4): 1603-1610.

[6] Tang J, Duan H, Lao S. Swarm intelligence algorithms for multiple unmanned aerial vehicles collaboration: A comprehensive review[J]. Artificial Intelligence Review, 2023, 56(5): 4295-4327.

[7] Schmuck, Patrik, Ziegler, et al. COVINS: Visual-Inertial SLAM for Centralized Collaboration[C]//2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2021). 2021.

[8] Hao Xu, Peize Liu, Xinyi Chen, et al. D^2SLAM: Decentralized and Distributed Collaborative Visual-Inertial SLAM System for Aerial Swarm[J]. IEEE Transactions on Robotics, 2024, Vol.40: 3445-3464.

[9] Liu, C., Zhao, J., Sun, N., 2022. A Review of Collaborative Air-Ground Robots Research. J. Intell. Robot. Syst. https://doi.org/10.1007/s10846-022-01756-4.

[10] Romanelli, F., Martinelli, F., Mattogno, S., 2023. Resilient Simultaneous Localization and Mapping Fusing Ultra Wide Band Range Measurements and Visual Odometry. J Intell Robot Syst 109, 64. https://doi.org/10.1007/s10846-023-01995-z.

[11] Du X, Ji C, Zhang L, et al. SP-VIO: Robust and Efficient Filter-Based Visual Inertial Odometry with State Transformation Model and Pose-Only Visual Description[J]. 2024.

[12] Wang, Liqiang, Tang, et al. PO-KF: A Pose-Only Representation-based Kalman Filter for Visual Inertial Odometry[J]. IEEE Internet of Things Journal,2025,: 1.

[13] Walter, Viktor, Staub, et al. UVDAR System for Visual Relative Localization with Application to Leader-Follower Formations of Multirotor UAVs(Article)[J]. IEEE Robotics and Automation Letters,2019,Vol.4(3): 2637-2644.

[14] Ge, R., Lee, et al. Vision-based Relative Detection and Tracking for Teams of Micro Aerial Vehicles[C]//2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2022.

[15] Schilling F, Schiano F, Floreano D (2021)Vision-Based Drone Flocking in Outdoor Environments. IEEE Robotics and Automation Letters 6(2):2954–2961. https://doi.org/10.1109/LRA.2021.3062298

[16] Vrba M, Saska M (2020) Marker-Less Micro Aerial Vehicle Detection and Localization Using Convolutional Neural Networks. IEEE Robotics and Automation Letters 5(2):2459–2466. https://doi.org/10.1109/LRA.2020.2972819

[17] Dubois R, Eudes A, Frémont V (2022) Sharing visual-inertial data for collaborative decentralized simultaneous localization and mapping. Robotics and Autonomous Systems 148:103933. https://doi.org/10.1016/j.robot.2021.103933

[18] Gross J, De Petrillo M, Beard J, et al (2019) Field-Testing of a UAV-UGV Team for GNSS-Denied Localization in Subterranean Environments. In: Proceedings of the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+), pp 2112–2124, https://doi.org/10.33012/2019.16912

[19] Zhu F, Ren Y, Kong F, et al (2023) Swarm-LIO: Decentralized Swarm LiDAR inertial Odometry. In: IEEE International Conference on Robotics and Automation (ICRA), pp 3254–3260, https://doi.org/10.1109/ICRA48891.2023.10161355

[20] Pritzl, V.,Vrba, et al. Fusion of Visual-Inertial Odometry with LiDAR Relative Localization for Cooperative Guidance of a Micro-Scale Aerial Vehicle [arXiv][J]. arXiv,2023.

[21] Nguyen TM, Cao M, Yuan S, et al (2022) VIRAL-Fusion: A Visual-Inertial-Ranging Lidar Sensor Fusion Approach. IEEE Transactions on Robotics 38(2):958-977. https://doi.org/10.1109/TRO.2021.3094157

[22] Delama, Giulio, Shamsfakhr, et al. UVIO: An UWB-Aided Visual-Inertial Odometry Framework with Bias-Compensated Anchors Initialization[C]//2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2023.

[23] P. -Y. Kao, H. -J. Chang, K. -W. Tseng, T. Chen, H. -L. Luo and Y. -P. Hung, "VIUNet: Deep Visual–Inertial–UWB Fusion for Indoor UAV Localization," in IEEE Access, vol. 11, pp. 61525-61534, 2023, doi: 10.1109/ACCESS.2023.3279292.

[24] Nguyen TH, Nguyen TM, Xie L (2022) Flexible and Resource-Efficient Multi-Robot Collaborative Visual-Inertial-Range Localization. IEEE Robotics and Automation Letters 7(2):928–935.

[25] Ziegler T, Karrer M, Schmuck P, et al (2021) Distributed Formation Estimation Via Pairwise Distance Measurements. IEEE Robotics and Automation Letters 6(2):3017–3024.

[26] Xu H, Wang L, Zhang Y, et al (2020) Decentralized Visual-Inertial-UWB Fusion for Relative State Estimation of Aerial Swarm. In: IEEE International Conference on Robotics and Automation (ICRA), https://doi.org/10.1109/ICRA40945.2020.9196944

[27] Cao Y, Beltrame G. VIR-SLAM: visual, inertial, and ranging SLAM for single and multi-robot systems. Auton Robot, 2021, 45(6):

[28] Zhang P, Chen G, Li Y, et al (2022) Agile Formation Control of Drone Flocking Enhanced With Active Vision-Based Relative Localization. IEEE Robotics and Automation Letters 7(3):6359–6366. https://doi.org/10.1109/LRA.2022.3171096

[29] Queralta JP, Li Q, Schiano F, et al (2022) VIO-UWB-Based Collaborative Localization and Dense Scene Reconstruction within Heterogeneous Multi-Robot Systems. In: International Conference on Advanced Robotics and Mechatronics (ICARM), https://doi.org/10.1109/ICARM54641.2022.9959470.

[30] Spasojevic I, Liu X, Ribeiro A, et al (2023) Active Collaborative Localization in Heterogeneous Robot Teams. In: Robotics: Science and Systems.

[31] Li, Jian, Yang, et al. Cooperative localization for UAVs in GNSS-denied area based on optimized belief propagation[J]. Measurement: Journal of the International Measurement Confederation, 2022, Vol. 192: 110797.

[32] Lin, Huei-Yung, Zhan, et al. GNSS-denied UAV indoor localization with UWB incorporated visual inertial odometry[J]. Measurement,2023, Vol.206: 112256.

[33] Xiong, Cheng, Lu, et al. Onboard cooperative relative positioning system for Micro-UAV swarm based on UWB/Vision/INS fusion through distributed

graph optimization[J]. Measurement: Journal of the International Measurement Confederation, 2024, Vol.234:

[34] Tang, Xiushan, Yang, et al. A collaborative localization algorithm for UAV Ad Hoc network based on improved sequence quadratic programming and unscented Kalman filtering in GNSS denied area[J]. Measurement, 2025, Vol.242: 115977.

[35] Qin T, Li P, Shen S. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. IEEE Trans Robot, 2018, 34(4): 1004～1020.

[36] Patrick Geneva; Kevin Eckenhoff; Woosik Lee;Yulin Yang; Guoquan Huang. OpenVINS: A Research Platform for Visual-Inertial Estimation[A].2020 IEEE International Conference on Robotics and Automation (ICRA)[C],2020.

[37] Q. Cai, Y. Wu, L. Zhang, and P. Zhang, "Equivalent constraints for two-view geometry: Pose solution/pure rotation identification and 3d reconstruction," Int. J. Comput. Vis., vol. 127, no. 2, pp. 163–180, FEB 2019.

[38] Q. Cai, L. Zhang, Y. Wu, W. Yu, and D. Hu, "A pose-only solution to visual reconstruction and localization," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 1, pp. 73–86, 2023.

[39] Wang, Liqiang, Tang, et al. PO-KF: A Pose-Only Representation-based Kalman Filter for Visual Inertial Odometry[J]. IEEE Internet of Things Journal,2025,: 1.

Note: Figure translations are in progress. See original paper for figures.