remote sensing

MDPI

Article

An Integrated Algorithm Fusing UWB Ranging Positioning and Visual–Inertial Information for Unmanned Vehicles

Shuang Li [1,2,3], Lihui Wang [1,*], Baoguo Yu [2,3], Xiaohu Liang [1,2,3], Shitong Du [2,3], Yifan Li [2,3] and Zihan Yang [2,3]

[1] School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China; 230209205@seu.edu.cn (S.L.); 230209200@seu.edu.cn (X.L.)
[2] State Key Laboratory of Satellite Navigation System and Equipment Technology, Shijiazhuang 050081, China; yubg@seu.edu.cn (B.Y.); dushitong@hrbeu.edu.cn (S.D.); liyifan1996@hrbeu.edu.cn (Y.L.); yangzh@nuaa.edu.cn (Z.Y.)
[3] The 54th Research Institute of China Electronics Technology Group Corporation, Shijiazhuang 050081, China
* Correspondence: wlhseu@seu.edu.cn

Abstract: During the execution of autonomous tasks within sheltered space environments, unmanned vehicles demand highly precise and seamless continuous positioning capabilities. While the existing visual–inertial-based positioning methods can provide accurate poses over short distances, they are prone to error accumulation. Conversely, radio-based positioning techniques could offer absolute position information, yet they encountered difficulties in sheltered space scenarios. Usually, three or more base stations were required for localization. To address these issues, a binocular vision/inertia/ultra-wideband (UWB) combined positioning method based on factor graph optimization was proposed. This approach incorporated UWB ranging and positioning information into the visual–inertia system. Based on a sliding window, the joint nonlinear optimization of multi-source data, including IMU measurements, visual features, as well as UWB ranging and positioning information, was accomplished. Relying on visual inertial odometry, this methodology enabled autonomous positioning without the prerequisite for prior scene knowledge. When UWB base stations were available in the environment, their distance measurements or positioning information could be employed to institute global pose constraints in combination with visual–inertial odometry data. Through the joint optimization of UWB distance or positioning measurements and visual–inertial odometry data, the proposed method precisely ascertained the vehicle's position and effectively mitigated accumulated errors. The experimental results indicated that the positioning error of the proposed method was reduced by 51.4% compared to the traditional method, thereby fulfilling the requirements for the precise autonomous navigation of unmanned vehicles in sheltered space.

Keywords: sheltered space; unmanned vehicle positioning; ultra-wideband ranging; visual–inertia odometry

check for updates

Citation: Li, S.; Wang, L.; Yu, B.; Liang, X.; Du, S.; Li, Y.; Yang, Z. An Integrated Algorithm Fusing UWB Ranging Positioning and Visual–Inertial Information for Unmanned Vehicles. *Remote Sens.* **2024**, *16*, 4530. https://doi.org/10.3390/rs16234530

Academic Editor: Baocheng Zhang

Received: 29 September 2024
Revised: 3 November 2024
Accepted: 5 November 2024
Published: 3 December 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In outdoor space, unmanned vehicles can utilize Beidou, the global position system (GPS), and other satellite navigation systems to accurately determine their positions. However, when these vehicles transition into indoor or underground environments, they encounter challenges in effectively receiving signals from Earth-orbiting navigation satellites due to signal attenuation caused by building shielding. As a result, achieving precise positioning in sheltered environments presents a substantial obstacle for the autonomous operation of unmanned vehicles [1–3].

Currently, a variety of research methodologies have been developed for localizing unmanned vehicles in sheltered space, including dead reckoning, visual positioning, and wireless positioning techniques. The dead reckoning method relies on the utilization of an inertial measurement unit (IMU) and vehicle odometer to compute both position and attitude. By measuring accelerometer and gyroscope data, we can determine the

acceleration and angular velocity of the unmanned vehicle. Through the integration of these data, real-time information regarding speed and attitude can be obtained autonomously without relying on external environmental cues. However, due to the inherent drift error issues within IMU systems, the accuracy gradually diminishes over time. Furthermore, odometry-based positioning methods may introduce significant errors as a result of sliding during motion [4,5].

The visual positioning methods can be divided into two categories: the single-camera positioning method and the visual odometer (VO) method. For the single-camera positioning method, when prior accurate information about the reference object is present, the camera of the unmanned vehicle is utilized to capture the fixed reference object, and the space rear intersection method is employed to determine its own pose. However, due to the limitations of camera image resolution, this method is only suitable for close-range positions, where direct observation of the reference object is feasible. When the distance is large, it exhibits significant positioning errors, and when the reference object cannot be observed, it fails [6,7]. The visual odometer method continuously captures images during the movement of the unmanned vehicle, extracting and matching features between key image frames. It constructs a geometric measurement model to assess the changes in feature point movements and subsequently estimates the posture variations of the unmanned vehicle. Since the visual odometer struggles with simple rotation tracking, it is commonly combined with inertial sensors in practical applications. After combination, the visual inertial odometer (VIO) can be stably applied to autonomously position unmanned vehicles without prior information. Although VIO can calculate precise positioning accuracy under favorable lighting conditions and with high-quality images, it essentially remains a relative positioning method that suffers from cumulative errors and cannot provide global absolute position information without other prior knowledge. Consequently, numerous studies have been conducted to enhance long-range visual localization by incorporating design elements such as global signposts or supplementary global information [8–11]. The principle of the radio positioning method refers to the technology of determining the position by analyzing the spatial propagation characteristics of radio signals. The positioning method can be primarily categorized into two approaches: the signal angle measurement method and the signal propagation time measurement method. In the signal angle measurement method, the direction of a radio signal is determined by receiving signals from various directions and making use of the directional characteristics of the antenna. The signal propagation time measurement method involves calculating the distance by measuring the time taken for a radio signal to travel from transmission to reception and then employing triangulation principles to determine precise positions [12,13]. Despite extensive research and application of radio technology in sheltered space positioning, in certain scenarios, the accuracy of distance measurements is reduced due to multipath effects and non-line-of-sight propagation. Moreover, radio-based positioning methods typically require the deployment of three or more base stations for signal transmission and demand prior knowledge of their locations, thus limiting practical applications.

Each positioning method has its own inherent limitations. Relying solely on a single information source is incapable of meeting the requirements of the increasingly complex application scenarios for unmanned vehicles. Integrated navigation combines multiple different single systems and makes use of multiple information sources to complement each other, thereby achieving more robust and accurate positioning capabilities. However, extensive data fusion also presents challenges in terms of computational complexity and high processor requirements. Therefore, it is of great significance to research multi-source data fusion using a limited number of radio signals to establish a positioning system with enhanced redundancy and accuracy [14]. In view of the limitations of the filter fusion method, a fusion method based on factor graph optimization is proposed to improve it, and the joint optimization objective function is established and solved through nonlinear optimization theory. The main contributions of this study are as follows:

(1) A sliding window model and nonlinear optimization theory are utilized to establish an objective optimization function. The optimal pose solution is obtained through iterative optimization, thereby forming a positioning system that integrates visual inertia with the ranging of a single base station of UWB.
(2) At present, most combinations merely utilize the ranging information of UWB without considering the angle constraint. In this study, the angle information acquired by combining the antenna array is employed to enhance the constraint of UWB.
(3) Simulated environments and real sheltered space environments are constructed for testing purposes. Experimental design is introduced, encompassing experimental data, experimental schemes, experimental results, and analysis. Through comparative analysis, it is observed that the algorithm proposed in this study exhibits superior performance in relevant indicators.

The remainder of this study is arranged as follows: Section 2 introduces the existing multi-source fusion localization technology and the commonly used models and methods. Section 3 presents the proposed strategy and the process of the algorithm. Section 4 introduces the experimental design, including the experimental data, experimental results, and data analyses. Section 5 concludes this study.

2. Related Work

The fundamental concept of multi-sensor information fusion involves weighted averaging, where different weights are assigned to the data collected by various sensors based on their respective accuracies. Multi-sensor fusion algorithms can be classified into filtering-based and nonlinear-based methods according to different fusion principles. Many traditional classical filtering algorithms, such as the particle filter, Kalman filter, hidden Markov algorithm, and dynamic Bayesian network, stem from Bayesian filtering. The core idea of the Bayesian filtering algorithm is to calculate the corresponding prior probability distribution and information fusion eigenvalue based on the known sensor input data, thereby describing the overall environmental characteristics provided by the prior model [15]. Nonlinear optimization-based algorithms include methods, like steepest descent method, Gauss–Newton algorithm, and LM algorithm. Multi-sensor fusion can be categorized into tight coupling and loose coupling, depending on the degree of integration. In loose coupling, each module is estimated separately before their positioning results are fused. In contrast, tight coupling involves processing and computing the original data collected by two or more sensors to obtain a positioning result. However, this approach brings about challenges related to noise interaction between different sensors. Consequently, implementing a tight coupling algorithm is more complex than a loose coupling one, but it offers more comprehensive sensor data fusion capabilities, leading to better results.

Hausman, K. et al. [16] added GPS, inertia, vision, UWB and other positioning modules to the UAV positioning system in a modular form. Their work mainly focuses on how to achieve self-alignment and initialization among these modules. Nyqvist, H. E. et al. [17] adopted a filtering method based on EKF, which fused UWB ranging information as observations with visual inertial odometer to obtain a global six-degrees-of-freedom pose estimation without drift. A self-built dataset was used for testing, the positioning error was better than 0.2 m, and the angle error was better than 6°. Xu, X. et al. [18] proposed a filtering method based on MCKF. This method first calculated the initial angle using the output of the three sensors, then fused the UWB and IMU data, and built its own dataset for testing. The positioning error was less than 0.4 m. Yang, B. et al. [19] proposed a fusion method based on UVIP, which first optimized binocular PL-VIO, combined the optimized pose results with UWB, constrained the distance and pose, and tested the self-built dataset. The horizontal positioning error was less than 0.15 m, and the rotation angle error was less than 4.5°. Shen, B. et al. [20] proposed a fusion method based on ES-EKF, which fused the VIO positioning results of the binocular S-MSKF algorithm with the UWB positioning results and then built its own dataset for testing. The horizontal positioning error was about 0.3 m. Yang, B. et al. [21] proposed a fusion method named R-UVIS, which improved

UVIP by incorporating the UWB distance residual into VIO for joint optimization and then set up its own dataset for testing. The horizontal positioning error was less than 0.13 m, and the rotation angle error was less than 3.3°. Cao, Y. et al. [22] proposed a fusion method named VIR-SLAM, which was applied to double-layer sliding window UWB, monocular and inertial odometer of a single robot, and the transformation estimation module for collaborative SLAM. A self-built dataset was used for testing, with a horizontal positioning error of less than 0.15 m. Juan, K. et al. [23] proposed a fusion method named U-VIO, which used UWB measurement for front-end pose estimation and back-end loop optimization. The UWB measurement values in previous key frames were used as loop closure constraints, and self-built datasets were used for testing. The positioning error was less than 0.2 m. Gao, B. et al. [24] proposed a fusion method named PL-VIO/UWB. In this method, PL-VIO odometer and UWB were tightly coupled for optimization, and self-built datasets were used for testing, with a horizontal positioning error of 0.12 m. Fan, Q. et al. [25] employed a magnetic anti-interference method to eliminate error values of the UWB sensor in non-line-of-sight environments and then filtered the inertial data. Subsequently, an adaptive Kalman filter was devised for enhanced accuracy in data fusion. Zhong, Y. et al. [26] proposed a fusion approach based on an extended Kalman filter that combined UWB sensor and inertial data, along with defining criteria for non-line-of-sight scenarios to ensure precise and continuous positioning information in complex environments. Wang, C. et al. [27] introduced a vision/inertia positioning system augmented by UWB technology, integrating visual tracking constraints into an attitude map using proposed smooth and virtual distance constraints to optimize robust trajectory estimation. Table 1 describes a brief overview of the latest related methods mentioned above.

Table 1. An overview of the latest related methods.

Author	Algorithm	Information Source Type	Positioning Error (m)
Nyqvist, H. E. et al. [17]	EKF	IMU, Vision, UWB	0.2
Xu, X. et al. [18]	MCKF	IMU, UWB	0.4
Yang, B. et al. [19]	UVIP	IMU, Vision, UWB	0.15
Shen, B. et al. [20]	ES-EKF	IMU, Vision, UWB	0.3
Yang, B. et al. [21]	R-UVIS	IMU, Vision, UWB	0.13
Cao, Y. et al. [22]	VIR-SLAM	IMU, Vision, UWB	0.15
Juan, K. et al. [23]	U-VIO	IMU, Vision, UWB	0.2
Gao, B. et al. [24]	PL-VIO/UWB	IMU, Vision, UWB	0.12
Fan, Q. et al. [25]	DSAKF	IMU, UWB	0.18
Wang, C. et al. [27]	graph optimization	IMU, vision, UWB	0.11

In view of the shortcomings of the traditional visual inertial odometer, which easily produces cumulative errors, this study proposes introducing a global sensor, UWB. Specifically, in the case of a single-base-station UWB, the UWB ranging value is integrated with the visual and inertial systems. The target optimization function is established through the sliding window model and nonlinear optimization theory, and the optimal pose solution is obtained through iterative optimization. Different from the aforementioned methods, this study effectively utilizes the ranging and angle measurement information of the UWB single base station. Additionally, pose optimization is achieved through a factor map, reducing the number of dependent base stations and attaining a plug-and-play effect.

3. Method

This section first establishes the technical framework of multi-source fusion positioning presented in this study. Then, it introduces the positioning principle of the UWB single

base station. After that, it presents the visual inertial odometry method. Finally, it explains the fusion positioning model of visual inertial odometry and UWB based on a factor graph.

3.1. Method Overview

In this study, a system architecture is proposed, which integrates UWB ranging and positioning information with visual inertial odometry to achieve precise positioning of unmanned vehicles within sheltered space. The model utilizes multiple sensor data sources, including UWB ranging, angle measurement, image processing, acceleration, and angular velocity measurements, thereby providing more accurate and reliable position estimation results. When using UWB information, although the single-station positioning range is limited, it can still offer distance information between the unmanned vehicle and the positioning base station by measuring arrival time, arrival angle, or signal strength. The visual inertial odometry primarily relies on camera and inertial navigation sensors. Through image processing technology, the camera extracts environmental feature points, which are then used to calculate motion change information based on their constraint relationships. The IMU captures attitude and position changes through the integration of acceleration and angular velocity measurements over time. By fusing the camera and IMU data, the motion trajectory of the unmanned vehicle can be obtained. While UWB distance measurement does not suffer from error accumulation issues, it may be affected by occlusion problems. Visual inertial positioning provides good continuity but may experience significant drift under low-texture conditions [28,29]. The core idea behind the multi-source ranging and vision–inertial fusion model architecture is to combine multiple sensor data sources to enhance the accuracy and reliability of position estimation and to design an adaptive fusion positioning strategy that addresses differences based on the distance between the unmanned vehicle and the radio base station. The proposed system architecture is shown in Figure 1.

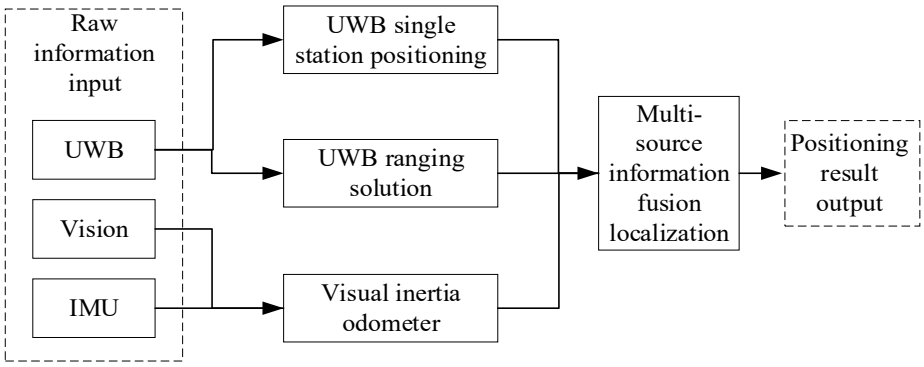


Figure 1. Architecture diagram of the multi-source fusion positioning model.

3.2. Principle of UWB Single-Base-Station Positioning

The UWB single-station positioning technology is different from the traditional method, which requires at least four anchor points to jointly determine the position. With only one UWB base station, three-dimensional positioning can be achieved. First, the symmetrical bilateral bidirectional ranging method is used to calculate the distance between the base station and the tag. Then, the pitch angle and horizontal angle are determined by measuring the phase angle of the array antenna. Finally, an equation is constructed based on the geometric relationship between distance and angle to calculate the three-dimensional position information of the tag [30]. The principle of symmetrical bilateral bidirectional ranging is shown in Figure 2.

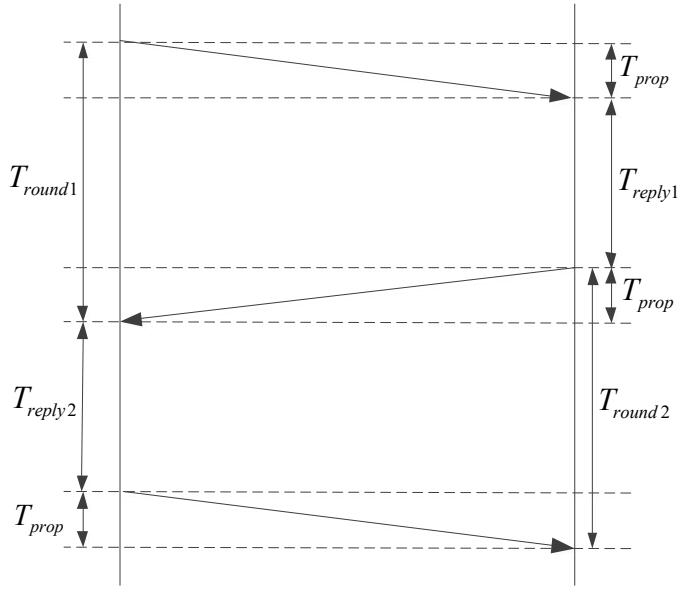


Figure 2. Principle of symmetrical bilateral bidirectional ranging method.

Upon receiving the response signal from the positioning tag, the positioning base station will initiate another transmission of a ranging message. Subsequently, this message is received by the positioning tag. Each device accurately records the timestamps for transmitting and receiving these message frames. By utilizing these data, it becomes feasible to calculate the round-trip delay of both signals.

$$\begin{aligned} T_{round1} &= 2T_{prop} + T_{reply1} \\ T_{round2} &= 2T_{prop} + T_{reply2} \end{aligned} \tag{1}$$

where T_{round1} and T_{round2} represent the time it takes a signal to travel back and forth between the positioning tag and base station, T_{reply1} and T_{reply2} represent the response time, T_{prop} represents the signal propagation time. Considering the whole process of information exchange, the following equations can be obtained:

$$T_{round1} + T_{round2} = 4T_{prop} + T_{reply1} + T_{reply2} \tag{2}$$

$$T_{prop} = \frac{\left(\left(T_{round1} - T_{reply2}\right) + \left(T_{round2} - T_{reply1}\right)\right)}{4} \tag{3}$$

T_{round1} and T_{round2}, T_{reply1} and T_{reply2} can be obtained from the timer of the positioning base station and the positioning tag, respectively. The radio propagation speed in air is approximately equal to the speed of light, and then the distance between the positioning base station and the tag can be obtained as follows:

$$d = T_{prop}c \tag{4}$$

where d represents the distance between the positioning base station and the tag; c represents the speed of light.

A uniform circular array consists of N array elements uniformly distributed on a circle with radius R, and its radius is $R \leq \lambda/2$; λ is the carrier wavelength, as shown in Figure 3. With the center of the circle as the origin of three-dimensional space coordinates, the coordinates of the matrix n of the uniform circular array are:

$$p_n = \left(R\cos(\frac{2\pi n}{N}), R\sin(\frac{2\pi n}{N}), 0\right), n \in [1, N] \tag{5}$$

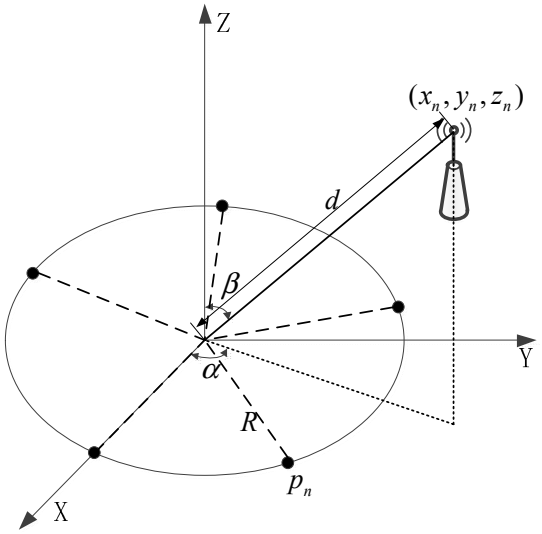


Figure 3. Schematic diagram of array antenna angle measurement.

Relative to the reference point circle point, the receiving signal delay of the array element n is:

$$\Delta t_n = -\frac{R\cos(\frac{2\pi n}{N})\sin\beta\cos\alpha + R\sin(\frac{2\pi n}{N})\sin\beta\sin\alpha}{c} \tag{6}$$

According to the trigonometric formula, the above formula can be simplified as:

$$\Delta t_n = -\frac{R}{c}\sin\beta\cos(\frac{2\pi n}{N} - \alpha) \tag{7}$$

Then, the phase difference between the matrix n and the reference point can be obtained as:

$$\Delta\phi_n(\beta,\alpha) = -\frac{2\pi R}{\lambda}\sin\beta\cos(\frac{2\pi n}{N} - \alpha) \tag{8}$$

According to the phase difference between multiple elements and reference points, the horizontal angle and the pitch angle can be obtained using the following formula. The following overdetermined equations can be solved by the least quadratic method.

$$\begin{bmatrix}\phi_{1-0}\\ \phi_{2-0}\\ \cdots\\ \phi_{n-0}\end{bmatrix} = \begin{bmatrix}-\frac{2\pi R}{\lambda}\sin\beta\cos(\frac{2\pi\cdot 1}{N} - \alpha)\\ -\frac{2\pi R}{\lambda}\sin\beta\cos(\frac{2\pi\cdot 2}{N} - \alpha)\\ \cdots\\ -\frac{2\pi R}{\lambda}\sin\beta\cos(\frac{2\pi n}{N} - \alpha)\end{bmatrix} \tag{9}$$

$$S(\alpha,\beta) = \min\sum_{i=1}^{n}(-\frac{2\pi R}{\lambda}\sin\beta\cos(\frac{2\pi i}{N} - \alpha) - \phi_{i-0})^2 \tag{10}$$

The principle of single-station positioning based on angle measurement and ranging is also illustrated in Figure 3. The geometric relationship among coordinates, angle information, and distance of the target label is represented in Equation (11).

$$(x_n, y_n, z_n) = (d\sin\beta\cos\alpha, d\sin\beta\sin\alpha, d\cos\beta) \tag{11}$$

where (x_n,y_n,z_n) represents the position of the UWB label, d represents the distance between the base station antenna and the positioning label, α and β represent the pitch angle and the horizontal angle, respectively. The position of the target label can be obtained by combining Formula (4), Formula (10) and Formula (11).

3.3. Principle of Visual Inertia Odometer

To optimize the state variables of the unmanned vehicle, the fusion of visual and inertial information requires the implementation of a sliding window technique. The specific formulation for optimizing the state quantities is as follows:

$$\begin{aligned} X &= [x_0, x_1, \ldots, x_n, l_0, l_1, \ldots, l_m] \\ x_k &= [p_k^w, v_k^w, q_k^w, b_{a_k}^b, b_{\omega_k}^b] \; k \in \psi[0, n] \end{aligned} \tag{12}$$

where X represents the number of states to be optimized, encompassing n unmanned vehicle states and m inverse depths associated with visual feature points. x_k represents the position p_k^w, speed v_k^w, attitude q_k^w, acceleration deviation $b_{a_k}^b$ and gyroscope deviation $b_{\omega_k}^b$ of the unmanned vehicle at the k key moment; the upper corner mark w represents the world coordinate system; the upper corner mark b represents the inertial coordinate system; l_m represents the inverse depth of a visual feature point.

$$X = \min\left\{ \sum_{k\in\psi} \left\| r_I\left(\hat{z}_{b_{k+1}}^{b_k}, X\right)\right\|^2 + \sum_{(l,j)\in c} \rho\left(\left\| r_c\left(\hat{z}_l^{c_j}, X\right)\right\|^2\right) \right\} \tag{13}$$

where $r_I\left(\hat{z}_{b_{k+1}}^{b_k}, X\right)$ represents the measurement residual of IMU pre-integration between two key frames in the sliding window, and $r_c\left(\hat{z}_l^{c_j}, X\right)$ represents the visual reprojection residual. $\hat{(\cdot)}$ represents the observation that contains noise. $\rho(\cdot)$ represents Huber kernel function. In the process of nonlinear optimization, if an abnormal observation value is generated due to the mismatching of features or other reasons, adding it to the graph optimization will introduce a large error edge and lead the system in the wrong optimization direction. Therefore, to increase the robustness of nonlinear optimization, the Huber kernel function is used in this study to reduce the weight of abnormal observations during the optimization process.

(1) IMU pre-integral residual term

The IMU pre-integration component between two consecutive keyframes can be expressed as follows:

$$\begin{aligned} \alpha_{b_{k+1}}^{b_k} &= \iint_{t\in[t_k,t_{k+1}]} R_t^{b_k}(\hat{a}_t - b_{a_t})dt^2 \\ \beta_{b_{k+1}}^{b_k} &= \int_{t\in[t_k,t_{k+1}]} R_t^{b_k}(\hat{a}_t - b_{a_t})dt \\ \gamma_{b_{k+1}}^{b_k} &= \int_{t\in[t_k,t_{k+1}]} \frac{1}{2}\Omega(\hat{\omega}_t - b_{\omega_t})\gamma_t^{b_k}dt \end{aligned} \tag{14}$$

where $\alpha_{b_{k+1}}^{b_k}$, $\beta_{b_{k+1}}^{b_k}$ and $\gamma_{b_{k+1}}^{b_k}$ represent the position change amount, velocity change amount and attitude change amount of inertial navigation relative to its own coordinate system, respectively. $\hat{a}_t$ and $\hat{\omega}_t$ represent the measurements of the accelerometer and gyroscope, respectively. $R_t^{b_k}$ represents the rotation matrix. b_{a_t} and b_{ω_t} represent the deviation of the accelerometer and gyroscope, respectively.

The IMU residual term is defined as the error between the position, speed, and rotation of two keyframes and the amount of change in accelerometer and gyroscope deviation:

$$r_I\left(\hat{z}_{b_{k+1}}^{b_k}, X\right) = \begin{bmatrix} \delta\alpha_{b_{k+1}}^{b_k} \\ \delta\beta_{b_{k+1}}^{b_k} \\ \delta\theta_{b_{k+1}}^{b_k} \\ \delta b_a \\ \delta b_\omega \end{bmatrix} = \begin{bmatrix} R_w^{b_k}(P_{b_{k+1}}^w - P_{b_k}^w + \frac{1}{2}g^w\Delta t_k^2 - v_{b_k}^w\Delta t_k) - \hat{\alpha}_{b_{k+1}}^{b_k} \\ R_w^{b_k}(v_{b_{k+1}}^w + g^w\Delta t_k - v_{b_k}^w) - \hat{\beta}_{b_{k+1}}^{b_k} \\ 2\left[q_{b_k}^{w-1}\otimes q_{b_{k+1}}^w\right]\otimes(\hat{\gamma}_{b_{k+1}}^{b_k})^{-1} \\ b_a^{b_{k+1}} - b_a^{b_k} \\ b_\omega^{b_{k+1}} - b_\omega^{b_k} \end{bmatrix} \tag{15}$$

(2) Visual residual term

Taking the feature point l continuously observed in the sliding window as an example, the construction of a visual residual term is introduced. It is assumed that the feature point l is observed by the images of frame i and frame j, respectively, and frame i is the key frame, which is observed for the first time in the sliding window. Based on binocular vision observation, three kinds of visual reprojection equations used to construct visual residuals in this study are introduced.

$$r_c\left(\hat{z}_l^{c_j}, X\right) = \begin{bmatrix} \delta P_l^{c_{j,1}} \\ \delta P_l^{c_{j,2}} \\ \delta P_l^{c_{i,2}} \end{bmatrix} = \begin{bmatrix} R_b^{c_1}(R_w^{b_j}(R_{b_i}^w(R_{c_1}^b \frac{1}{\lambda_l}\pi_c^{-1}(\begin{bmatrix} \hat{u}_l^{c_{i,1}} \\ \hat{v}_l^{c_{i,1}} \end{bmatrix}) + p_{c_1}^b) + p_{b_i}^w - p_{b_j}^w) - p_{c_1}^b) - \hat{P}_l^{c_{j,1}} \\ R_b^{c_2}(R_w^{b_j}(R_{b_i}^w(R_{c_1}^b \frac{1}{\lambda_l}\pi_c^{-1}(\begin{bmatrix} \hat{u}_l^{c_{i,1}} \\ \hat{v}_l^{c_{i,1}} \end{bmatrix}) + p_{c_1}^b) + p_{b_i}^w - p_{b_j}^w) - p_{c_2}^b) - \hat{P}_l^{c_{j,2}} \\ R_b^{c_2}((R_{c_1}^b \frac{1}{\lambda_l}\pi_c^{-1}(\begin{bmatrix} \hat{u}_l^{c_{i,1}} \\ \hat{v}_l^{c_{i,1}} \end{bmatrix}) + p_{c_1}^b) - p_{c_2}^b) - \hat{P}_l^{c_{i,2}} \end{bmatrix} \quad (16)$$

where $\left[\hat{u}_l^{c_{i,1}}, \hat{v}_l^{c_{i,1}}\right]$ represents the pixel coordinate of the feature point l observed for the first time in the sliding window; $c_{i,1}$ represents the left-eye image of frame i; π_c^{-1} represents the projection matrix represented by parameters in the camera, used to project pixel coordinates to a normalized coordinate system; $(R_{c_1}^b, p_{c_1}^b)$ and $(R_{c_2}^b, p_{c_2}^b)$ represent the external parameters of the left camera and the right camera with respect to the IMU, respectively; $P_l^{c_{j,1}}$ and $P_l^{c_{j,2}}$ represent the reprojection results of feature point l projected from the left camera image of frame i to the left and right camera images of frame j; $P_l^{c_{i,2}}$ represents the result of reprojection from the left camera to the right camera of the frame i. After the above visual observation equation is obtained, the residual term can be obtained by subtracting the observed value from the calculated value.

3.4. Visual Inertial Odometer and UWB Fusion Positioning Model Based on Factor Map

The filter-based approach presupposes a Markov property, meaning that the state at time k is only related to the state at time $k-1$, and is not related to the state and observations prior to the time $k-1$. However, in the actual system, the current frame is often related to several previous frames, making the filter-based method more difficult to solve. Moreover, there is no outlier detection mechanism in the filter-based method, which causes the system to have difficulty converging and be more prone to divergence when there are outliers. In visual localization and UWB localization, outliers are very common. Therefore, this work presents an optimized sliding window structure that integrates visual inertia information and UWB information, including the comprehensive advantages of these two types of data. Specifically, it utilizes the high relative pose accuracy provided by visual inertia fusion estimation and incorporates the global information constraints offered by UWB ranging and positioning. The optimized variables of the system encompass measurements from visual inertial odometry, UWB distance, and UWB positioning [31,32]. Figure 4 illustrates the relationship between visual inertial odometry and UWB ranging information.

Here, $\mathrm{p}(x_A, y_A)$ represents the momentary coordinate of the unmanned vehicle at position point A, (x_B, y_B) represents the momentary coordinate at position point B, (x_O, y_O) represents the position coordinate of the UWB positioning station.

The equation that establishes the relationship between the unmanned vehicle's location coordinates and the results obtained from radio fusion ranging and visual inertial positioning can be derived by considering the positional arrangement of the unmanned vehicle with respect to the UWB positioning station depicted in Figure 4.

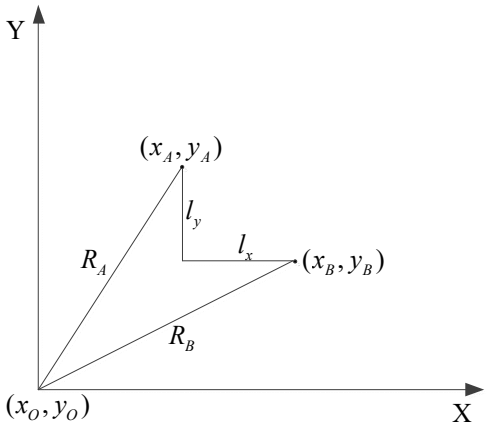


Figure 4. The spatial relationship between the unmanned vehicle and the UWB positioning station.

$$(x_A - x_o)^2 + (y_A - y_o)^2 = R_A^2 \tag{17}$$

$$(x_B - x_o)^2 + (y_B - y_o)^2 = R_B^2 \tag{18}$$

where $R_A = (t_A - t_{oA})c$ represent the fusion ranging result from the unmanned vehicle to the base station at time A, and $R_B = (t_B - t_{oB})c$ represent the fusion ranging result from the unmanned vehicle to the base station at time B. The position coordinates of A and B satisfy the following relation:

$$x_A = x_B - l_x \tag{19}$$

$$y_A = y_B - l_y \tag{20}$$

The estimated poses obtained from VIO and the UWB ranging results are utilized as local poses, while the UWB station positioning results serve as global poses. These poses are then converted into factor terms within the factor map and subsequently inputted into the global pose map to construct optimization problems. Ultimately, this process yields locally accurate and globally drift-free six-degrees-of-freedom pose estimation. The structure of the global factor diagram is presented as follows.

The nodes of the factor map, depicted in Figure 5, represent poses in the world coordinate system. The green edge connecting two consecutive nodes signifies a local pose constraint derived from visual inertial odometry. Similarly, the yellow edge represents a local constraint obtained through UWB ranging information, while the blue edge denotes a global constraint derived from UWB positioning information as a global sensor.

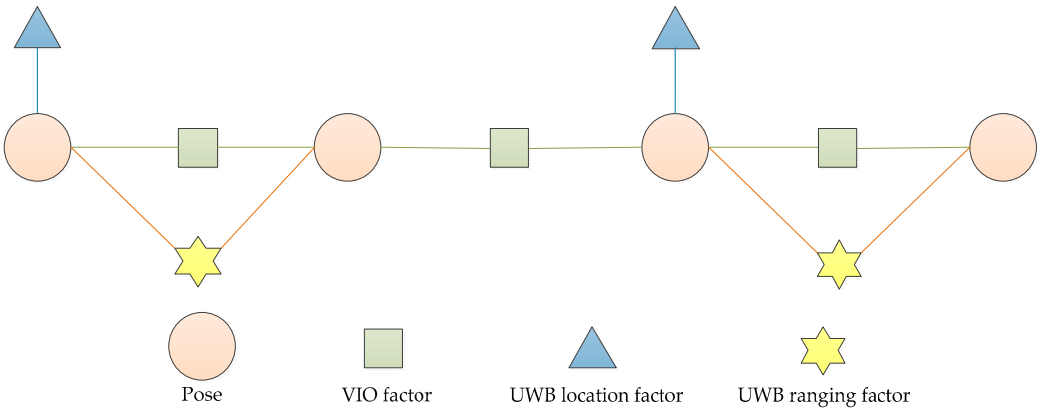


Figure 5. UWB positioning/ranging and visual–inertial fusion based on factor map.

The factor graph optimization problem can be regarded as a maximum likelihood estimation task, where each pose node corresponds to an optimization variable. Nonlinear

optimization, based on a sliding window approach, $\mathrm{X} = [\mathrm{x}_0, \mathrm{x}_1, \ldots, \mathrm{x}_n], \mathrm{x}_i = [\mathrm{p}_i^w, \mathrm{q}_i^w]$, is employed to extract pose information from the closest n key frames within the window at any given moment. The nonlinear optimization problem can be effectively tackled using either the Gauss–Newton method or the Levenberg–Marquardt method.

$$\mathrm{X} = \operatorname{argmin} \sum_{i=0}^{n} \left\{ \| z_i^{\mathrm{VIO}} - h_i^{\mathrm{VIO}}(\mathrm{X}_i - \mathrm{X}_{i-1}) \|^2 + \| z_i^{\mathrm{UWB}} - h_i^{\mathrm{UWB}}(\mathrm{X}_i - \mathrm{X}_{i-1}) \|^2 + \| z_i^{\mathrm{UWB}_d} - h_i^{\mathrm{UWB}_d}(\mathrm{X}_i - \mathrm{X}_{i-1}) \|^2 \right\} \quad (21)$$

where $z_i^{\mathrm{VIO}} - h_i^{\mathrm{VIO}}(\mathrm{x}_i - \mathrm{x}_{i-1})$ is the residual of visual inertia odometer, $z_i^{\mathrm{UWB}} - h_i^{\mathrm{UWB}}(\mathrm{x}_i - \mathrm{x}_{i-1})$ is the residual of UWB single station positioning result, and $z_i^{\mathrm{UWB}_d} - h_i^{\mathrm{UWB}_d}(\mathrm{x}_i - \mathrm{x}_{i-1})$ is the UWB ranging result parameter.

Since VIO can provide a very accurate estimate in a short amount of time, the VIO optimization factor is used to construct the visual inertial odometry residual term using the relative pose between two successive frames obtained by VIO. Define the error between the observed UWB coordinate value and the predicted value as the UWB factor term, where the predicted value is the coordinates obtained by VIO.

$$z_i^{\mathrm{UWB}} - h_i^{\mathrm{UWB}}(\mathrm{x}_i - \mathrm{x}_{i-1}) = p_i^{\mathrm{UWB}} - p_i^{\mathrm{VIO}} \quad (22)$$

The UWB positioning value imposes constraints on the node's position, while the covariance matrix calculated during UWB positioning represents the covariance of this factor term. Residuals between the UWB tag and the base station are constructed to enhance estimation accuracy. We define the status value as the distance value calculated based on VIO positioning results and the base station, while considering the observed value as directly measured by UWB. Here, we define the residual term as the difference between the status distance value and observed value.

$$z_i^{\mathrm{UWB}_d} - h_i^{\mathrm{UWB}_d}(\mathrm{x}_i - \mathrm{x}_{i-1}) = z_i^{\mathrm{UWB}_d} - \| p_i^{\mathrm{VIO}} - p_i^{\mathrm{A}} \| \quad (23)$$

where p_i^{A} represents the coordinates of the base station.

4. Experiments

To assess the proposed system, two experiments were carried out. In the first experiment, the proposed algorithm was compared with two other advanced algorithms within a simulated environment. The second experiment involved comparing and analyzing the performance of these three algorithms in a real-world environment. Subsequently, a numerical analysis was conducted to demonstrate the effectiveness of our method.

4.1. Test and Analysis in Simulation Environment

This section first introduces the establishment of the simulation platform, then gives the evaluation method of positioning accuracy, and finally conducts experimental tests in the simulation environment to verify the positioning performance of the algorithm.

4.1.1. Simulation Experimental Environment Setup

The absence of standardized datasets and open-source algorithms poses a challenge for comparing and analyzing the accuracy of different algorithms. To validate the proposed method in this study, a sheltered space environment model was initially constructed using Gazebo 9.12.0 simulation software. Simultaneously, an unmanned vehicle equipped with a camera, inertia sensors, and other sensors was simulated. Subsequently, Wireless Inset 3.4.4 simulation software was employed to simulate the UWB signal propagation characteristics. The obtained simulation results were then integrated into the UWB single-station location algorithm. Finally, MATLAB 2022b was utilized to simulate the system's positioning capability.

The architecture of the simulation system for sheltered space positioning consists of three component modules: (1) the simulation module for the sheltered space environment and unmanned vehicle sensors; (2) the simulation module for radio propagation characteristics; (3) the simulation module for multi-source information fusion positioning. The architecture of the simulation system for sheltered space positioning is illustrated in Figure 6.

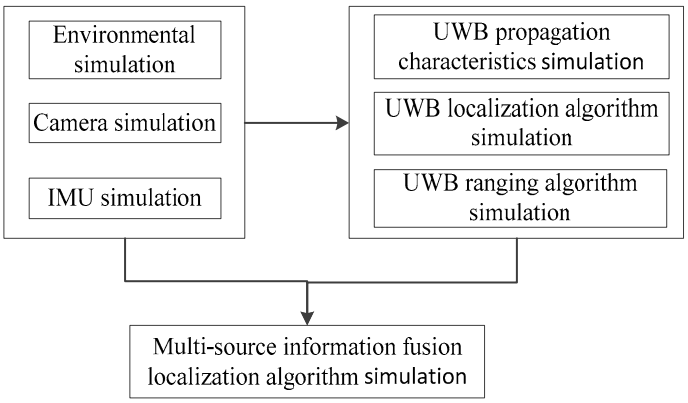


Figure 6. The architecture of the simulation system for sheltered space positioning.

The functions of each component module in the system are as follows: (a) The simulation module for sheltered space environment and unmanned vehicle sensors is responsible for highly realistic simulations of synthetic surface terrain, light, shadow, and background. It focuses on accurately simulating the configuration of the unmanned vehicles, including its structure appearance, wheel suspension, and wheel friction. In terms of sensor simulation, it aims to accurately simulate camera output image size parameters, IMU data, and other related parameters. (b) The simulation module for radio propagation characteristics primarily analyzes UWB signal propagation attenuation, multipath benefits, and ranging results through comprehensive and detailed simulations. (c) The simulation module for multi-source information fusion positioning specifically focuses on integrating UWB positioning results with ranging results and visual inertial odometer results using a reliable fusion algorithm. Users can define the geographical environment in the simulation system, while real-time data from various simulated sensors can be published.

4.1.2. Performance Metrics

The root-mean-square error (RMSE) is computed in Equation (24) to facilitate the comparison of positioning performance among different methods.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum\left((x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2\right)^2} \tag{24}$$

where (x_i, y_i) represents the real position of the test point I; $(\hat{x}_i, \hat{y}_i)$ represents the position obtained by different methods for the corresponding test point.

4.1.3. Simulation Experimental Tests and Analysis

Simulation parameter setting was as follows: The environment was constructed with dimensions of 100 m × 100 m. The camera's release frequency was set at 30 Hz, with a resolution of 680 × 480 pixels. The inertial sensor had a release frequency of 200 Hz and exhibited bias stability at an accuracy level of 0.8°/h. For movement within the environment, the unmanned vehicle's maximum speed was limited to 3 m/s and it could change its heading angle at a maximum rate of 3°/s. Starting from the origin point, the unmanned vehicle followed a predetermined moving route. To evaluate our method and compare it with other state-of-the-art localization techniques, the methods in [20,21] were

chosen for comparison. Figure 7 illustrates the trajectory obtained through the different methods, with the red line representing the actual path of the unmanned vehicle.

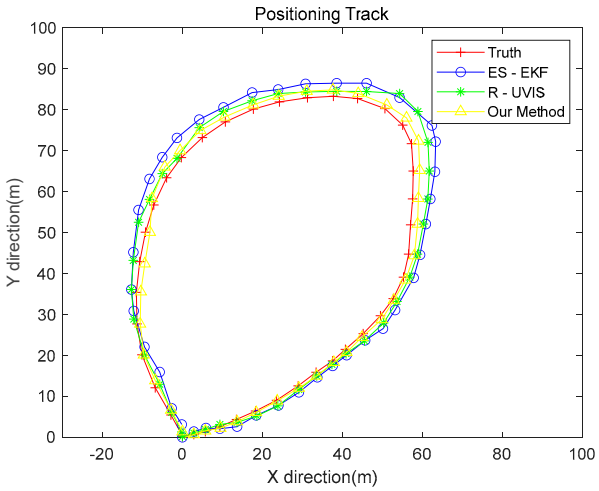


Figure 7. Trajectory diagram of various positioning methods in simulation environment.

As depicted in Figure 7, the average error of all positioning methods gradually increased as the distance between the unmanned vehicle's location and the base station expanded. However, by integrating UWB ranging and positioning information, our method demonstrated reduced positioning errors and effectively mitigated divergence. Consequently, enhanced precision in determining the position of the unmanned vehicle was achieved through the combination of UWB ranging and positioning information.

The size and cumulative distribution of positioning errors are shown in Figure 8a–c, respectively.

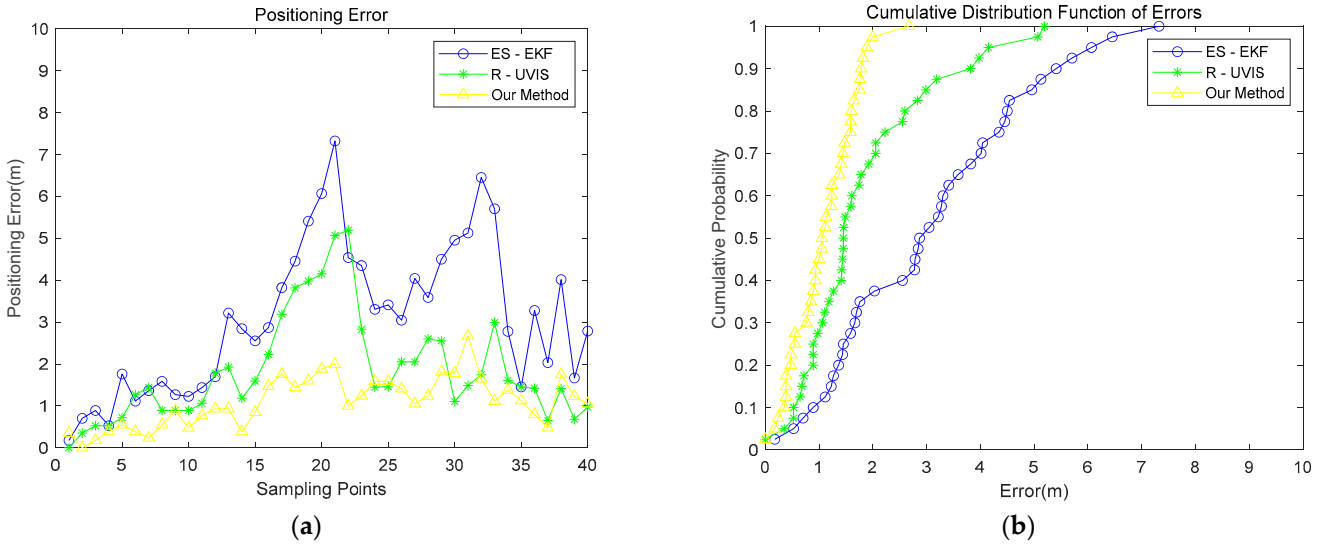
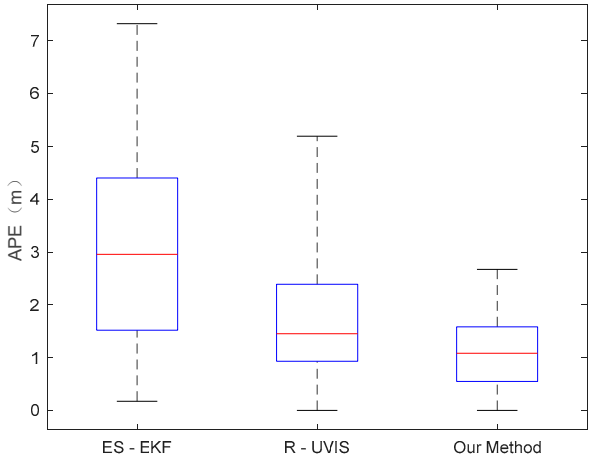


Figure 8. *Cont.*



(**c**)

Figure 8. Statistical results of fusion positioning errors: (**a**) positioning error diagram; (**b**) positioning error cumulative distribution diagram; (**c**) positioning error box plot.

The data in Table 2 show that the maximum error of our method was 2.68 m and the mean error was 1.10 m, with 95% of the errors being better than 2.5 m. Comparative experiments demonstrated that our method had higher localization accuracy than other methods.

Table 2. Statistics on the fusion positioning errors.

Experiment	Algorithm	Maximum (m)	Mean (m)	Median (m)	Root Mean Square Error (m)
Simulation scene	ES-EKF	7.32	3.08	2.96	3.54
	R-UVIS	5.19	1.83	1.45	2.20
	Our Method	2.68	1.10	1.08	1.25

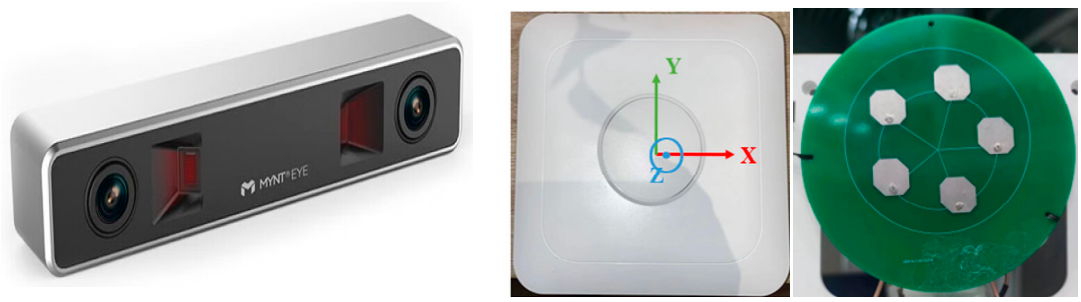
4.2. Test and Analysis in Real Environment

This section initially presents the real test environment and experimental equipment. Subsequently, the test was conducted in the actual scene. Finally, a statistical analysis of the test results was carried out to verify the positioning performance of the algorithm in the real environment.

4.2.1. Real Experimental Environment Setup

To validate the positioning capability of the proposed method in a real sheltered space, an indoor scene was selected for experimental verification. The test and validation equipment mainly consists of an unmanned ground vehicle (UGV), UWB positioning base station, and total station. Among them, the unmanned vehicle was developed by 54 Institute of China Electronics Technology Group Corporation and was equipped with a binocular camera, UWB positioning tag, and industrial computer. The static positioning accuracy of the UWB base station was 10 cm, and the dynamic positioning accuracy reached 30 cm in the unobstructed environment. The industrial computer utilized an i7-1165G7 CPU with 16 GB memory and ran on the Ubuntu18.04 OS. The RTK measuring reference point was located outside the test site with known latitude and longitude coordinates. Considering that hardware synchronization had a high cost and the speed of the unmanned vehicle in this study was relatively low, with a low requirement for synchronization accuracy, software synchronization was adopted. The Robot Operating System provided synchronization classes to complete multi-sensor time synchronization. In the test environment, total stations were used to measure the latitude and longitude coordinates of any given

points. The binocular camera and UWB positioning base station are shown in Figure 9a,b, respectively. Some indicators of the binocular camera, IMU, and UWB base stations are shown in Table 3.



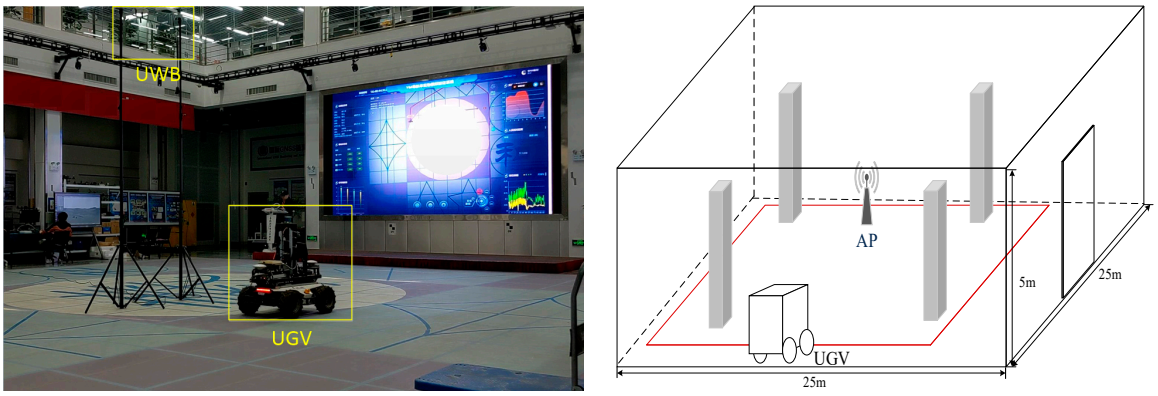
(**a**) (**b**)

Figure 9. Equipment related to experimental testing: (**a**) binocular camera; (**b**) UWB base station.

Table 3. Sensor performance parameters.

Device Model	Performance	Parameter Value
MYNTEYE	frequency	30 FPS
	resolution	752 × 480
	field of view Angle	D:146° H:122° V:76°
BMI088 IMU	frequency	200 Hz
	axis	6 axis
UWB	radio frequency	6~9 GHz
	number of antenna array	5
	static positioning accuracy	10 cm
	dynamic positioning accuracy	30 cm
	angular estimation accuracy	3°

The indoor scene positioning test was conducted at the artificial intelligence navigation test field of 54 Institute of China Electrical Science and Technology in Shijiazhuang city. The test site measured 25 m by 25 m with a height of 5 m, and the autonomous vehicle followed a predetermined path within a red rectangular boundary on the test site. The UWB base station was located at the center of the red rectangular enclosure, with an antenna elevation of 5 m, as illustrated in Figure 10.



(**a**) (**b**)

Figure 10. Indoor test scenario: (**a**) indoor test site diagram; (**b**) indoor three-dimensional structure diagram.

4.2.2. Real Experimental Tests and Analysis

Two experiments were conducted at the test site, where the movement direction and speed of the unmanned vehicle were remotely controlled. Each experiment involved a minimum distance traveled of 30 m, and the resulting trajectories are depicted in Figure 11.

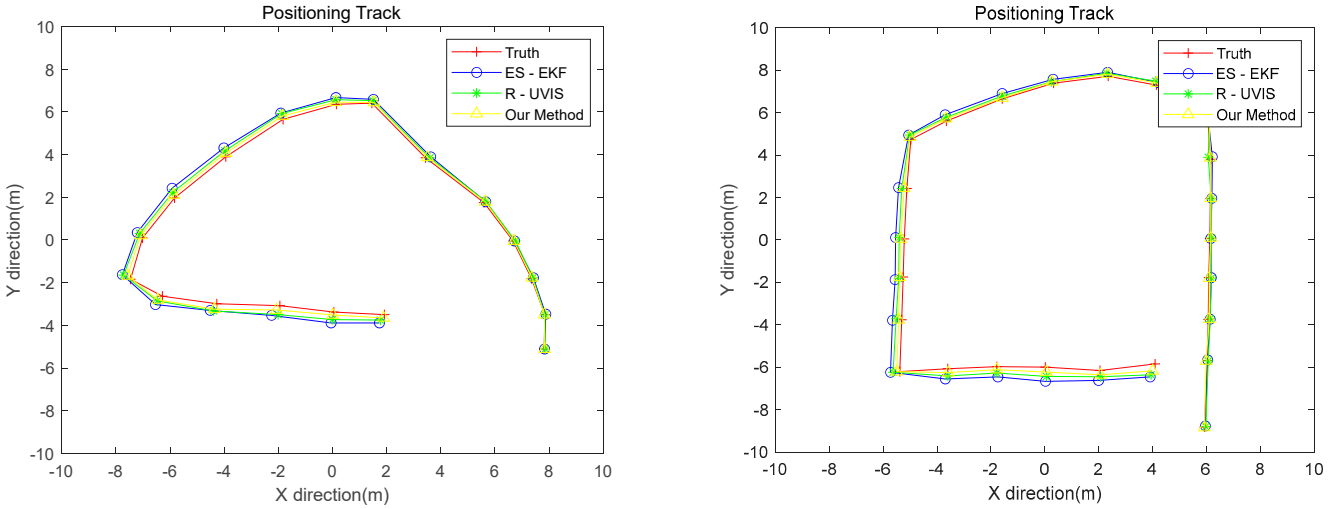


Figure 11. Moving trajectory diagram: (**a**) test results of test I; (**b**) test results of test II.

The latitude and longitude coordinates of the systematically chosen reference point were measured using a total station at approximately 2 m intervals from the starting point to the end point. Subsequently, an unmanned vehicle initiated its journey from the starting point, traversed these selected reference points sequentially, recorded its own coordinate position at each stop, and calculated the deviation from the actual coordinates of each reference point. The experimental results are presented in Figure 12 below.

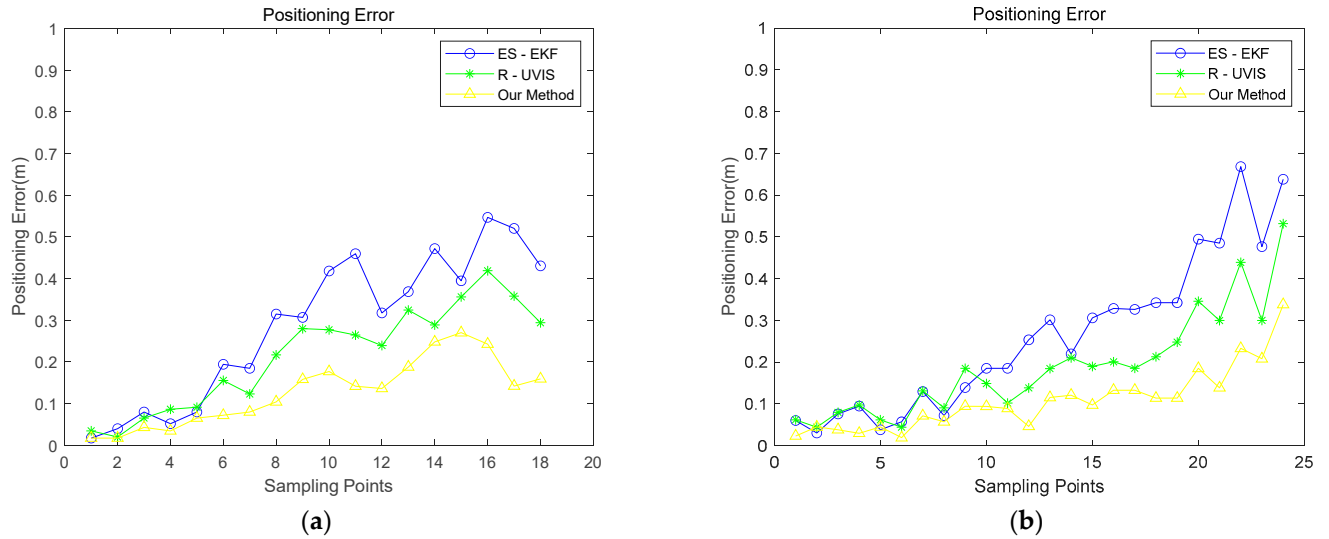


Figure 12. Statistical diagram of positioning error: (**a**) positioning error diagram of test I; (**b**) positioning error diagram of test II.

The cumulative distribution of positioning errors for both tests is illustrated in Figure 13.

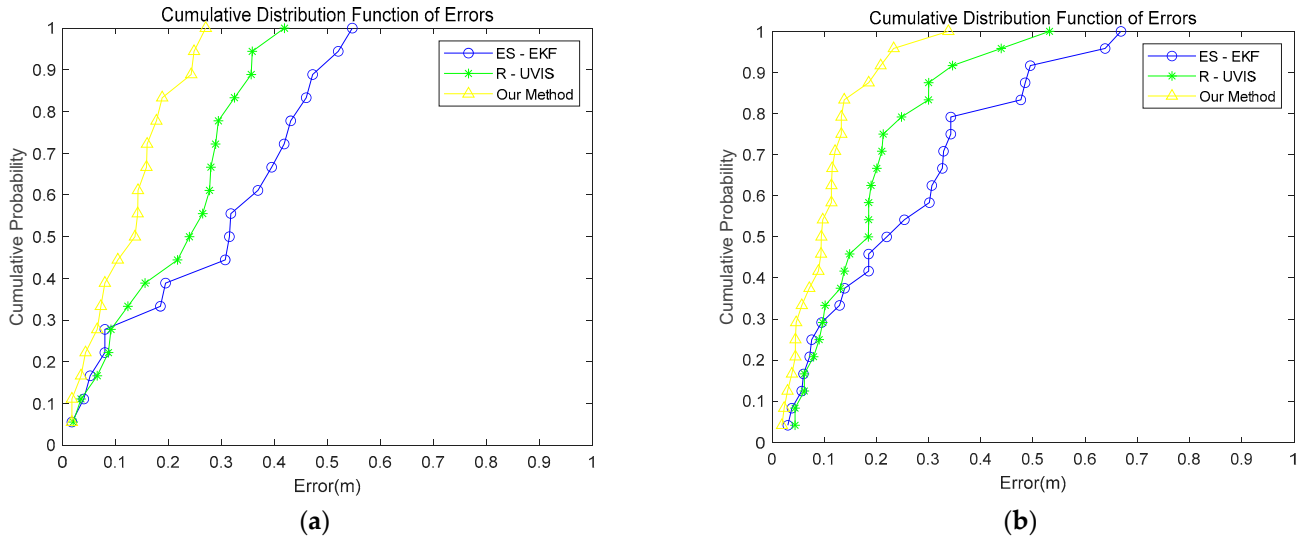


Figure 13. Cumulative distribution of positioning errors: (**a**) cumulative distribution diagram of test I; (**b**) cumulative distribution diagram of test II.

A comprehensive comparison was conducted among the ES-EKF method, the R-UVIS method, and our method, with the findings presented in Figure 14.

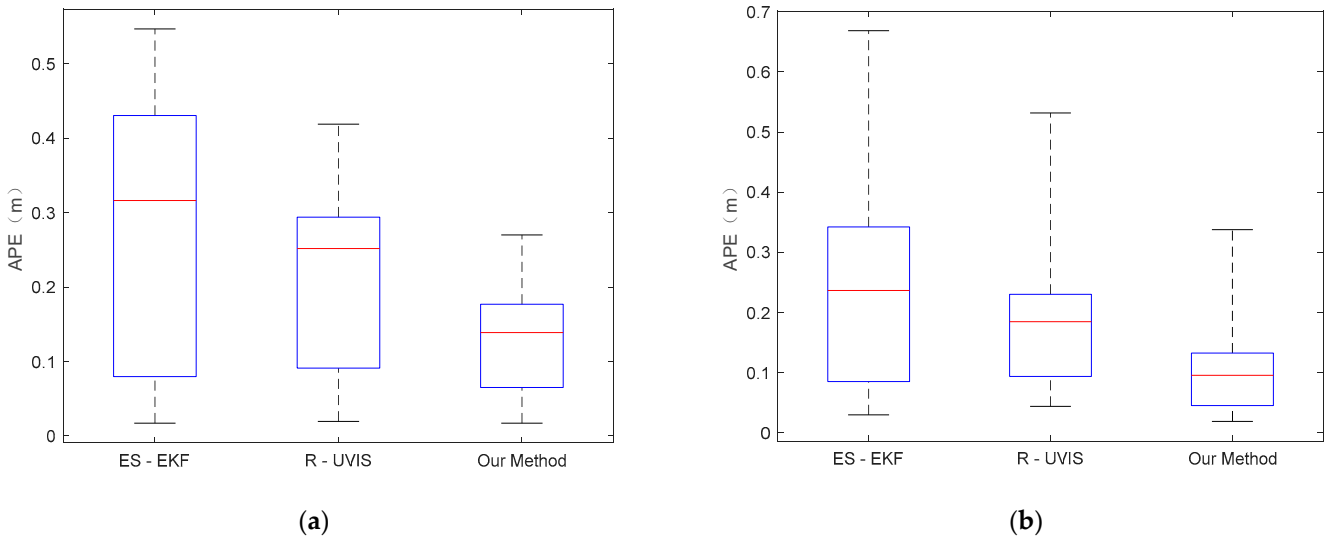


Figure 14. Comprehensive comparison of positioning errors: (**a**) positioning error results of indoor scene test I; (**b**) positioning error results of indoor scene test II.

The data in Table 4 show that the mean maximum error of our method was 0.30 m and the mean error was 0.12 m, with 95% of the errors being better than 0.25 m. The experimental results convincingly demonstrated that the VIO/UWB combination algorithm based on a factor graph had more accurate and reliable positioning performance. In conclusion, although UWB ranging and direction-finding information increased the dimension of the state vector and the complexity of the observation equation, they could also improve the system's ability to estimate the state. These additional constraints helped optimize the fusion algorithm to make better use of various information sources and improve the accuracy and stability of the system's positioning in complex environments.

Table 4. Statistics on the fusion positioning errors.

Experiment	Algorithm	Maximum (m)	Mean (m)	Median (m)	Root Mean Square Error (m)
Indoor scene test I	ES-EKF	0.55	0.29	0.32	0.34
	R-UVIS	0.42	0.22	0.25	0.25
	Our Method	0.27	0.13	0.14	0.15
Indoor scene test II	ES-EKF	0.67	0.26	0.23	0.32
	R-UVIS	0.53	0.19	0.18	0.22
	Our Method	0.33	0.11	0.09	0.13

5. Conclusions

To address the challenge of positioning unmanned vehicles within sheltered space, where prior environmental information is limited and radio-positioning base stations are in short supply, this work proposed a fusion-based positioning approach that combined UWB single-station distance measurements with visual inertial odometry. A plug-and-play elastic fusion framework based on factor graphs was designed to seamlessly integrate sensor data from multiple sources. The experimental results demonstrated that in the simulation environment, the maximum positioning error of this method was less than 2.68 m, the root mean square error was under 1.25 m, and the 2σ positioning error was less than 2.5 m. In the real environment, the experimental results indicated that the maximum positioning error of this method was less than 0.30 m, the root mean square error was less than 0.14 m, and the 2σ positioning error was less than 0.25 m.

In addition to focusing on multi-sensor fusion of UWB, vision, and IMU, the proposed framework can also be adapted for integrating pseudo-satellite, Bluetooth, 5G, and visual inertial odometer technologies. The test results revealed that as the unmanned vehicle moved farther away from the UWB base station, trajectory errors increased substantially, highlighting the limitations of relying solely on the UWB ranging values from a single base station. Future endeavors will explore a fusion positioning model that includes ranging constraints from multiple base stations, aiming to provide precise positioning for unmanned vehicles in sheltered space.

Author Contributions: All authors developed the idea that led to this study together. Conceptualization, L.W. and S.L.; methodology, S.L.; data processing and experimental analysis, X.L., S.D., Y.L., Z.Y. and S.L.; article writing and typesetting, S.L.; financial support, B.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Plan of China (No. 2021YFB3900803)) and supported by Hebei Natural Science Foundation (No. F2024523004).

Data Availability Statement: The original contributions presented in the study are included in the article material, and any further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yu, B.G.; Huang, L.; Bao, Y.; Jia, H.; Li, S. Research Status and Trends of Indoor Positioning and Navigation Technology in China. *J. Geod. Geoinf. Sci.* **2023**, *6*.
2. Li, X.; Wang, X.; Liao, J.; Li, X.; Li, S.; Lyu, H. Semi-tightly coupled integration of multi-GNSS PPP and S-VINS for precise positioning in GNSS-challenged environments. *Satell. Navig.* **2021**, *2*, 1. [CrossRef]
3. Liang, X.; Pan, S.; Yu, B.; Li, S.; Du, S. A Pseudo-Satellite Fingerprint Localization Method Based on Discriminative Deep Belief Networks. *Remote Sens.* **2024**, *16*, 1430. [CrossRef]
4. Liang, Y.; Müller, S.; Schwendner, D.; Rolle, D.; Ganesch, D.; Schaffer, I. A scalable framework for robust vehicle state estimation with a fusion of a low-cost IMU, the GNSS, radar, a camera and lidar. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October 2020–24 January 2021; pp. 1661–1668.
5. Xu, X.C. Lunar rover positioning and precision analysis for single closed station based on image. *Sci. Surv. Mapp.* **2015**.

6. Wang, L.; Jiang, Q.; Li, S. Research on Fusion Matching Object Recognition and Location Based on Binocular Image Multi-feature Point Fusion. *Radio Eng.* **2018**, *48*, 628–633.
7. Li, S.; Yu, B.; Jin, Y.; Huang, L.; Zhang, H.; Liang, X. Image-Based Indoor Localization Using Smartphone Camera. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 3279059. [CrossRef]
8. Liao, J.; Li, X.; Wang, X.; Li, S.; Wang, H. Enhancing navigation performance through visual-inertial odometry in GNSS-degraded environment. *Gps Solut.* **2021**, *25*, 50. [CrossRef]
9. Leutenegger, S.; Lynen, S.; Bosse, M.; Siegwart, R.; Furgale, P. Keyframe-b ased visual–inertial odometry using nonlinear optimization. *Int. J. Robot. Res.* **2015**, *34*, 314–334. [CrossRef]
10. Qin, T.; Li, P.; Shen, S. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Trans. Robot.* **2018**, *34*, 1004–1020. [CrossRef]
11. Ren, G.; Cao, Z.; Liu, X.; Tan, M.; Yu, J. Plj-slam: Monocular visual slam with points, lines, and junctions of coplanar lines. *IEEE Sens. J.* **2022**, *22*, 15465–15476. [CrossRef]
12. Cao, S.; Lu, X.; Shen, S. GVINS: Tightly coupled GNSS-visual-inertial fusion for smooth and consistent state estimation. *IEEE Trans. Robot.* **2022**, *38*, 2004–2021. [CrossRef]
13. Guo, Y.; Vouch, O.; Zocca, S.; Minetto, A.; Dovis, F. Enhanced EKF-based time calibration for GNSS/UWB tight integration. *IEEE Sens. J.* **2022**, *23*, 552–566. [CrossRef]
14. Shan, T.; Englot, B.; Ratti, C.; Rus, D. Lvi-sam: Tightly-coupled lidar-visual-inertial odometry via smoothing and mapping. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; pp. 5692–5698.
15. Qin, T.; Cao, S.; Pan, J.; Shen, S. A general optimization-based framework for global pose estimation with multiple sensors. *arXiv* **2019**, arXiv:1901.03642.
16. Hausman, K.; Weiss, S.; Brockers, R.; Matthies, L.; Sukhatme, G.S. Self-calibrating multi-sensor fusion with probabilistic measurement validation for seamless sensor switching on a UAV. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 4289–4296.
17. Nyqvist, H.E.; Skoglund, M.A.; Hendeby, G.; Gustafsson, F. Pose estimation using monocular vision and inertial sensors aided with ultra wide band. In Proceedings of the 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Banff, AB, Canada, 13–16 October 2015; pp. 1–10.
18. Xu, X.; Liu, X.; Zhao, B.; Yang, B. An extensible positioning system for locating mobile robots in unfamiliar environments. *Sensors* **2019**, *19*, 4025. [CrossRef]
19. Yang, B.; Li, J.; Zhang, H. Uvip: Robust uwb aided visual-inertial positioning system for complex indoor environments. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; pp. 5454–5460.
20. Shen, B.; Zhang, Z.; Shu, S. UWB-VIO integrated indoor positioning algorithm for mobile robots. *J. Comput. Appl.* **2022**, *42*, 3924.
21. Yang, B.; Li, J.; Zhang, H. Resilient indoor localization system based on UWB and visual–inertial sensors for complex environments. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 8504014. [CrossRef]
22. Cao, Y.; Beltrame, G. VIR-SLAM: Visual, inertial, and ranging SLAM for single and multi-robot systems. *Auton. Robot.* **2021**, *45*, 905–917. [CrossRef]
23. Jung, K.; Shin, S.; Myung, H. U-VIO: Tightly coupled UWB visual inertial odometry for robust localization. In Proceedings of the International Conference on Robot Intelligence Technology and Applications, Daejeon, South Korea, 16–17 December 2021; pp. 272–283.
24. Gao, B.; Lian, B.; Wang, D.; Tang, C. Low drift visual inertial odometry with UWB aided for indoor localization. *IET Commun.* **2022**, *16*, 1083–1093. [CrossRef]
25. Fan, Q.; Sun, B.; Sun, Y.; Zhuang, X. Performance enhancement of MEMS-based INS/UWB integration for indoor navigation applications. *IEEE Sens. J.* **2017**, *17*, 3116–3130. [CrossRef]
26. Zhong, Y.; Liu, T.; Li, B.; Yang, L.; Lou, L. Integration of UWB and IMU for precise and continuous indoor positioning. In Proceedings of the 2018 Ubiquitous Positioning, Indoor Navigation and Location-Based Services (UPINLBS), Wuhan, China, 22–23 March 2018; pp. 1–5.
27. Wang, C.; Zhang, H.; Nguyen, T.M.; Xie, L. Ultra-wideband aided fast localization and mapping system. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 1602–1609.
28. Liu, F.; Zhang, J.; Wang, J.; Han, H.; Yang, D. An UWB/vision fusion scheme for determining pedestrians' indoor location. *Sensors* **2020**, *20*, 1139. [CrossRef]
29. Taponecco, L.; D'Amico, A.A.; Mengali, U. Joint TOA and AOA estimation for UWB localization applications. *IEEE Trans. Wirel. Commun.* **2011**, *10*, 2207–2217. [CrossRef]
30. Nguyen, T.M.; Nguyen, T.H.; Cao, M.; Qiu, Z.; Xie, L. Integrated uwb-vision approach for autonomous docking of uavs in gps-denied environments. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 9603–9609.

31. Xu, H.; Wang, L.; Zhang, Y.; Qiu, K.; Shen, S. Decentralized visual-inertial-UWB fusion for relative state estimation of aerial swarm. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 8776–8782.
32. Li, S.; Yu, B.G.; Yi, Q.W.; Du, S.T.; Li, Y.F.; Yang, Z.H. A lunar rover location method based on the fusion of multi-source ranging and visual inertia. *Space Electron. Technol.* **2024**, *21*, 1–6.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.