

Article

Position Estimation Method for Small Drones Based on the Fusion of Multisource, Multimodal Data and Digital Twins

Shaochun Qu, Jian Cui, Zijian Cao, Yongxing Qiao, Xuemeng Men and Yanfang Fu *

School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710021, China; qushaochun@xatu.edu.cn (S.Q.); cuijian@st.xatu.edu.cn (J.C.); caozijian@xatu.edu.cn (Z.C.); qiaoyongxing@xatu.edu.cn (Y.Q.); menxuemeng@st.xatu.edu.cn (X.M.)

* Correspondence: fuyanfang@xatu.edu.cn

Abstract: In response to the issue of low positioning accuracy and insufficient robustness in small UAVs (unmanned aerial vehicle) caused by sensor noise and cumulative motion errors during flight in complex environments, this paper proposes a multisource, multimodal data fusion method. Initially, it employs a multimodal data fusion of various sensors, including GPS (global positioning system), an IMU (inertial measurement unit), and visual sensors, to complement the strengths and weaknesses of each hardware component, thereby mitigating motion errors to enhance accuracy. To mitigate the impact of sudden changes in sensor data, a high-fidelity UAV model is established in the digital twin based on the real UAV parameters, providing a robust reference for data fusion. By utilizing the extended Kalman filter algorithm, it fuses data from both the real UAV and its digital twin, and the filtered positional information is fed back into the control system of the real UAV. This enables the real-time correction of UAV positional deviations caused by sensor noise and environmental disturbances. The multisource, multimodal fusion Kalman filter method proposed in this paper significantly improves the positioning accuracy of UAVs in complex scenarios and the overall stability of the system. This method holds significant value in maintaining high-precision positioning in variable environments and has important practical implications for enhancing UAV navigation and application efficiency.

Keywords: UAV; multisource, multimodal data fusion; digital twins; position estimation



Citation: Qu, S.; Cui, J.; Cao, Z.; Qiao, Y.; Men, X.; Fu, Y. Position Estimation Method for Small Drones Based on the Fusion of Multisource, Multimodal Data and Digital Twins. *Electronics* **2024**, *13*, 2218. <https://doi.org/10.3390/electronics13112218>

Academic Editor: Franco Cicirelli

Received: 30 April 2024

Revised: 29 May 2024

Accepted: 4 June 2024

Published: 6 June 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

With the gradual opening of low-altitude airspace, the application of drones is becoming increasingly widespread in various fields, ranging from circuit inspection and logistics delivery to urban traffic management [1]. The effective deployment of drones has become a hallmark of contemporary technological progress; however, precise positioning and navigation capabilities are essential for drones to perform tasks safely and efficiently. Currently, there are multiple positioning methods for drones, including the global positioning system (GPS), inertial navigation system (INS), terrain-aided navigation (TAN), Doppler velocity Log (DVL), odometer (OD), and visual navigation (VO) [2,3]. Nevertheless, as application scenarios become more complex, the accuracy and robustness of navigation systems face severe challenges during high-speed maneuvers or prolonged flights.

Traditional drone positioning relies on a single sensor system, which may be adequate in specific usage environments, but in complex environments, the interference from multipath effects, signal blockage, and environmental noise can lead to the accumulation of sensor noise and motion errors, impacting positioning accuracy. For instance, the inertial navigation system (INS) is a type of relative positioning technology that can only provide accurate solutions for a limited time [4], as errors in inertial sensors and integration drift can lead to solution divergence [5,6]; therefore, to enhance the positioning accuracy and system stability of drones, researchers have begun to explore multisource and multimodal data fusion methods.

Data fusion technology [7] involves the merging and processing of multiple complementary or similar information sources, where the merged information is more precise than any single source. The information fusion technique of combined navigation involves first converting the navigation data from various sensors into a unified coordinate system, and then using appropriate mathematical estimation methods to fuse the data within this system, obtaining the system's optimal estimate [8,9]. Thus, the data fusion process entails using mathematical optimal estimation algorithms to estimate the status of the navigation system. There are several categories of data fusion algorithms, among which, analytical-based methods and learning-based methods are the most widely used. Analytical-based methods use analytical functions to simulate system states and external measurements, referred to as "estimation" in the literature [10–12]. Many analytical-based methods are widely utilized, such as Kalman filtering (KF) and particle filtering (PF) [13], while several learning-based data fusion methods, such as artificial neural networks (ANNs), fuzzy logic, and support vector machines (SVMs), are also used because they can model systems without prior statistical information about the process and measurement noise [14]. Since the introduction of the Kalman filter (KF), it has been extensively used to address many challenging and complex information fusion problems; however, traditional Kalman filtering can only be applied in linear systems. To apply federated Kalman filtering in nonlinear systems, several improved KF methods have been proposed [15]. Based on the changing Bayesian adaptive Kalman filtering [16], a method has been proposed to track the continuously varying noise variance in the global positioning system/inertial navigation system. Insufficient prior information can cause the filter to diverge [17]. The extended Kalman filter (EKF) is a widely used method for integrating the inertial/global positioning systems. It linearizes nonlinear systems through a first-order Taylor expansion [18]; for example, the fusion of visual sensors and an IMU can be divided into loose and tight coupling [19]. In loose coupling, an extended Kalman filter or the unscented Kalman filter (UKF) fuses the pose information obtained from the visual sensors and IMU as separate modules. The information fusion used in studies [20] is of loose coupling. The author utilizes tight coupling, where the pose information obtained after processing the intermediate data from the visual and IMU sensors through filters has higher accuracy compared to loose coupling, but its robustness is poorer, and the failure of a sensor can paralyze the entire system. Additionally, for multisensor data fusion, the time synchronization of data directly affects the accuracy of the fusion. A method of sequential measurement updating has been proposed in the literature [21], where the update timing nodes are the same as the IMU data measurement nodes. Other sensors act as auxiliary sensors to correct IMU data, and their update nodes must be on the IMU update nodes. Due to the different data acquisition frequencies of various sensors, it is necessary to synchronize the timing of various data at the fusion center, which can determine the best sequence for fusion [22]. The pre-fusion processes add many complex processing and computational procedures to the multisensor data fusion process.

Digital twin technology, as an innovative technology that has emerged in recent years, has demonstrated its significant application potential and value across multiple domains. For instance, in military operations [23], study reflects on the construction of a digital twin battlefield, proposing a universal architecture for digital twin battlefields and providing feasible solutions for their construction. In drone-related fields, study introduces an algorithm based on random finite sets, analyzing the system architecture and model, and demonstrating the application of digital twin robots in drone operations [24]. Study discusses and designs the digital twin framework for military large-scale drones using digital twin technology combined with cloud computing, analyzing its functions to some extent and providing solutions for constructing drone digital twin systems. In the area of drone positioning and navigation research, the introduction of digital twin technology offers new ideas and methods with which to address the challenges faced by traditional positioning systems. Digital twin technology creates a virtual drone model capable of real-time mapping and simulating the actual drone's state and behavior [25]. This bridge

between the virtual realm and reality not only aids in predicting and analyzing drone performance but also provides more precise positioning and navigation support under complex environmental conditions.

However, although multisource, multimodal data fusion and digital twin technology theoretically hold significant potential, effectively integrating these technologies to achieve higher positioning accuracy and robustness remains a challenge in practical applications. Most existing studies focus on improving specific technologies or algorithms, lacking an in-depth exploration of how to effectively merge these different technologies to meet the positioning needs of drones in complex environments.

Therefore, this paper proposes a novel positioning estimation method for small unmanned aerial vehicles (UAVs) based on multisource, multimodal data fusion and digital twin technology. As shown in Figure 1, this method initially utilizes various sensors, including GPS, IMU, and visual sensors, to acquire multimodal data. The IMU performs quaternion calculations to obtain attitude data, the RealSense camera provides environmental depth information as well as distance positioning, and the GPS performs coordinate system conversion in addition to providing global positioning data. These data are combined through tightly coupled optimization, integrating visual and inertial navigation data. The GPS fusion, based on pose graph optimization, is further refined, and the digital twin UAV is used as a critical data source. The extended Kalman filter (EKF) algorithm is employed to achieve data fusion between the real UAV and its digital twin.

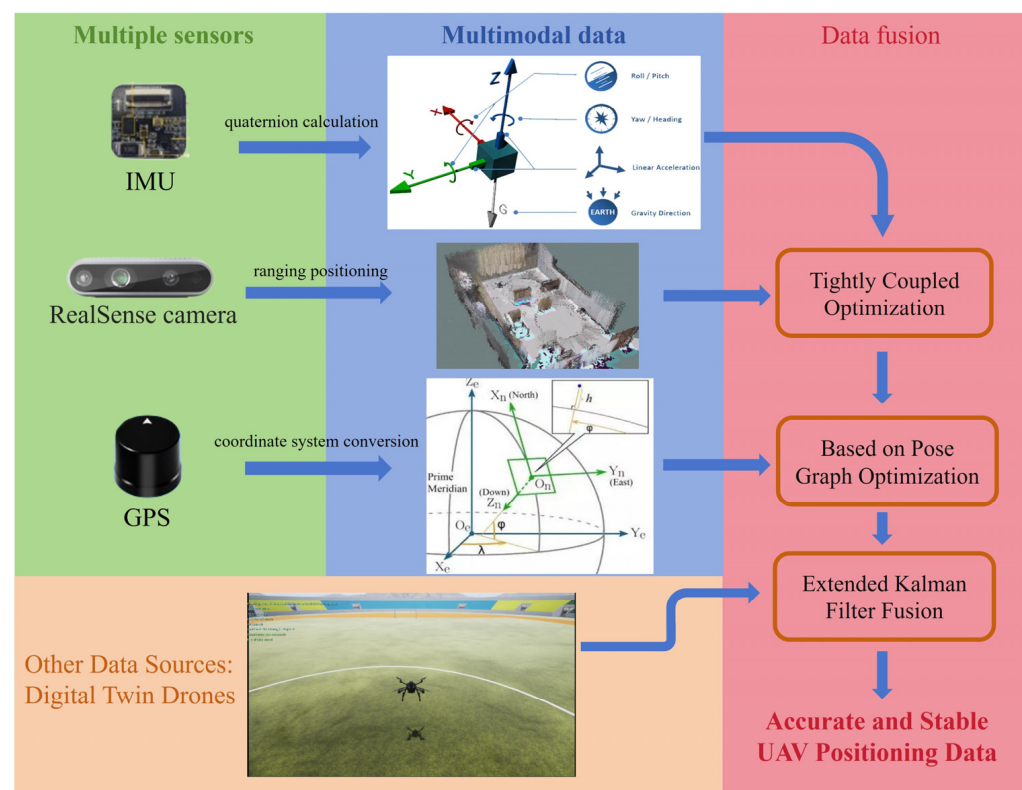


Figure 1. Schematic diagram of the multisource, multimodal data fusion positioning method.

This fusion approach effectively eliminates the limitations of a single sensor and reduces the positioning deviations caused by sensor errors as well as external environmental changes. Moreover, the digital twin drone provides a simulation model that predicts the real drone under the same conditions, utilizing the predictive capabilities of the digital twin model to correct and optimize the sensor data of the real drone. Especially in cases where sensor data are unstable or severely interfered with, it provides a more accurate reference for real-time location information, allowing for the timely correction of positioning

information when sensor data undergo sudden changes, ensuring the stability and accuracy of the drone control system. This study not only offers new ideas and methods for the precise positioning of small drones but also explores new possibilities for the practical application of digital twin technology in drone navigation and applications.

2. System and Model Construction

This paper establishes a data fusion system architecture based on digital twins for multisource and multimodal data fusion methods, as illustrated in Figure 2. This diagram depicts a multisource and multimodal data fusion system primarily deployed in UAV software layers and digital twin simulations. During the digital twin simulation process, sensor models provide the UAV flight control with sensor information obtained from the simulation environment. The flight controller computes the current state estimates and outputs actuation signals to the UAV model, which calculates the required forces as well as torques and assigns motor voltages. Additionally, inputs such as drag, friction, gravity, and torque are used by the physical engine to compute the UAV's next motion state in the simulation environment, and the estimated data are relayed in real time to the physical UAV. The UAV dynamics information, enriched with environmental model data like gravity, air density, atmospheric pressure, magnetic fields, and GPS coordinates, is then transmitted back to the sensor model, creating a feedback loop that captures sensor-acquired data. On the UAV's physical side, the system encompasses data collection, processing, and algorithmic fusion, utilizing raw data from the IMUs and GPS, along with environmental information captured by depth cameras. It also receives real-time location estimates from the digital twin UAV, which are fed into the data fusion module. The data fusion and control execution section form the core of the system. Here, data from various sensors are integrated to enhance positioning accuracy and robustness. The fused information is used to generate precise flight control signals, which are subsequently transmitted back to the UAV flight control through the MavLink control protocol, achieving accurate flight control.

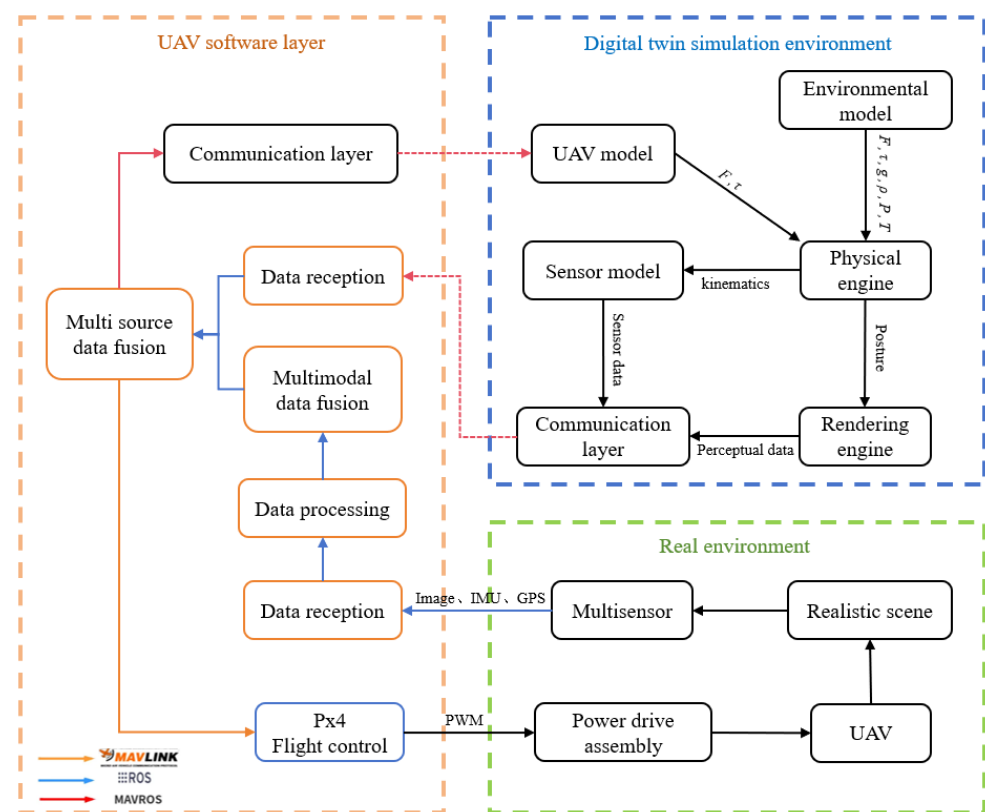


Figure 2. Architecture of the multisource and multimodal data fusion system.

Through this architecture, the UAV can swiftly adapt to complex environments and effectively handle uncertainties, enhancing its operational stability and reliability in complex application scenarios. Additionally, the architecture supports modularity and distributed processing, offering flexibility and scalability for the further development of UAV technology.

To effectively simulate and predict the behavior as well as performance of physical unmanned aerial vehicles (UAVs) and achieve consistency between the digital twin UAV and the physical UAV, it is essential to construct a multidimensional and multiscale UAV model within the digital twin framework. This foundation supports the data interaction between the virtual and physical UAVs. This article will provide detailed explanations of the model parameters and model composition, and then construct a multidimensional and multiscale UAV model from three aspects: geometric, physical, and behavioral. This enables the digital twin UAV to simulate the physical properties, dynamic behaviors, and environmental interactions of the physical UAV in real time. Through such modeling, the digital twin acts as a critical tool in aligning the virtual operations with the real-world dynamics of UAVs, enhancing predictive maintenance and operational planning.

2.1. Geometric Model

Incorporating the actual geometric dimensions and mass parameters of the unmanned aerial vehicle (UAV), a foundational model is developed using modeling software, based on the hardware architecture of the UAV as illustrated in Figure 3. Particular emphasis is placed on the construction of critical components such as motors, propellers, flight controllers, and GPS modules, which are integral to the flight control stability of the UAV. Detailed attention is devoted to the assembly specifications of each UAV component, as demonstrated in Figure 4, which depicts the UAV's external features, part dimensions, and material types, among other fundamental indicators. This process ensures a high-fidelity construction model of the physical UAV, enabling the digital twin system to accurately replicate the real UAV. This underpins subsequent digital twin UAV flight experiments by providing a fundamental model support, mitigating disparities between flight experimental results and actual conditions that may arise due to differences in geometric model parameters. Furthermore, it facilitates the development and testing of interfaces between the virtual and physical domains within digital twin technology.



Figure 3. Real drone and geometric model of the drone.

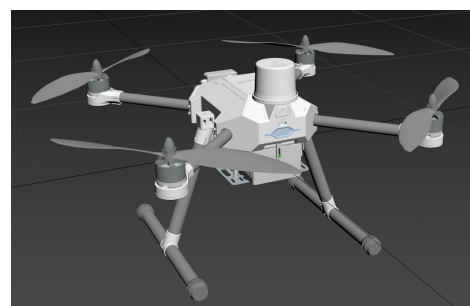


Figure 4. Digital twin drone model.

2.2. Physical Model

The virtual drone entity model primarily consists of two parts: built-in sensors of the PX4 flight controller, and physical connections. The built-in sensors mainly include a gyroscope, magnetometer, GPS, and barometer. The physical connection methods define the basic parameters for serial ports, local UDP, local TCP, remote TCP, QGC ground stations, and messaging protocols. The sensor module mainly contains devices such as gyroscopes, magnetometers, GPS, and barometers. The physical engine module mainly implements the dynamics modeling of quadrotor drones. A quadrotor drone with mass (m) is defined as a collection of four vertices $[x_1, x_2, x_3, x_4]$, each vertex having a control input $[u_1, u_2, u_3, u_4]$, where x_i represents the position of the vertex relative to the centroid and $u_i \in [0, 1]$ represents the corresponding control command. Finally, the control inputs, (u_i), are mapped to the forces, $F_i = f(u_i)$, and torques, $\tau_i = g(u_i)$, at vertex i . The digital twin UAV control input, u_i , is mapped to the angular velocity at each propeller located at the four vertices, and the local forces and torques at the four vertices are represented by Equations (1) and (2):

$$F_i = C_T \sigma \omega_{\max}^2 D^4 u_i \quad (1)$$

$$\tau_i = \frac{1}{2\pi} C_{pow} \sigma \omega_{\max}^2 D^5 u_i \quad (2)$$

wherein C_T and C_{pow} represent the thrust coefficient and power coefficient calculated based on the physical characteristics of the propeller, respectively. σ stands for air density, D is the propeller diameter, and ω_{\max} is the maximum angular velocity (rotations per second). Based on the parameters of the physical drone and the experimental environment, this paper assumes the related parameters of the drone, as shown in Table 1.

Table 1. Drone-related parameters.

Variable Name	Value
Thrust coefficient, C_T	0.109919
Pressure, σ	1.225 kg/m ³
Maximum angular velocity, ω_{\max}	106.61112 r/s
Propeller diameter, D	0.2286 m
Power coefficient, C_{pow}	0.040164

2.3. Behavior Model

The behavioral model represents a precise control framework for UAV flight, and a PID (proportional, integral, and derivative) control loop is an effective method for regulating UAV flight. Building upon this, the paper proposes a design for a dual closed-loop cascaded PID control circuit, as shown in Figure 5. In the diagram, the input signals, after filtration, are separately directed to the position PID controller and the attitude PID controller. The position PID controller takes the UAV's altitude value signal, which is linked to the UAV's NED (north, east, down) coordinate system. The signal represents the difference between the set target altitude and the actual flight altitude, with the controller's output being the vertical acceleration. The attitude PID controller receives the UAV's horizontal displacement signals, which are related to its horizontal coordinates. Its values represent the discrepancies in the UAV's roll, pitch, and yaw angles from the set horizontal target values to their actual values, with the controller's output being the horizontal acceleration. Through position and attitude calculations, the system converts the input quantities allocated to each direction into PWM (pulse width modulation) control signals, which are then sent to the motors. This process controls the motors, ensuring the precision of the UAV's flight control.

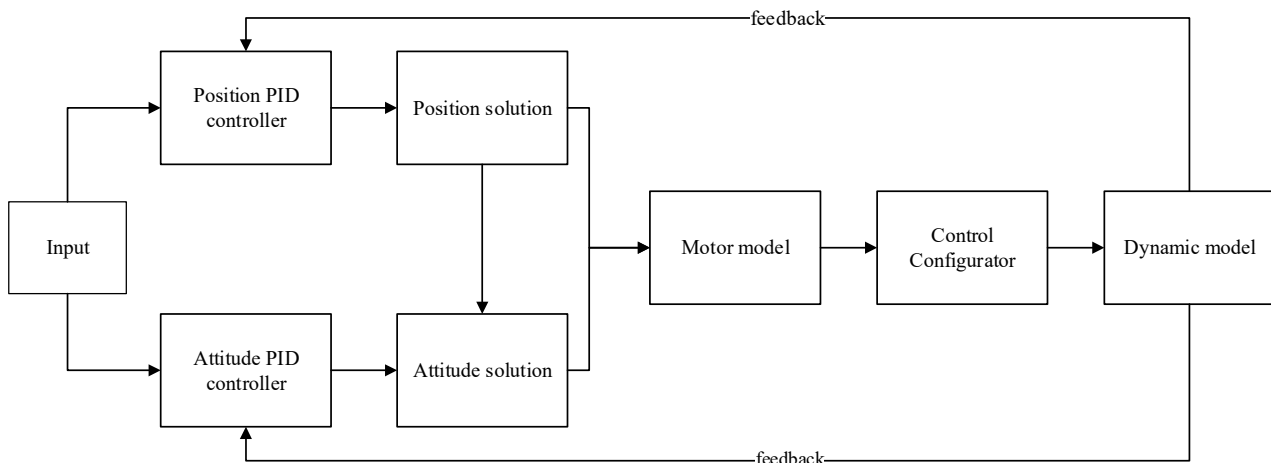


Figure 5. Double closed-loop cascade PID control loop.

3. Multisource, Multimodal Data Fusion

Navigating with only one type of sensor has its drawbacks. Due to the inherent differences among individual sensors, relying on a single sensor for navigation may lead to unsuitability for the environment, resulting in poor navigation accuracy. To address various complex environments, this paper adopts a navigation method using multiple sensors. The data collected by multiple sensors need to be fused, filtering out unnecessary data and extracting the required data, to achieve the best results.

3.1. Vision IMU Pose Fusion

Since pure visual positioning algorithms rely entirely on data obtained from cameras, they are susceptible to environmental lighting interference, resulting in low robustness. Cameras mounted on drones are prone to positioning failures during high-speed movements or rotational motions, making pure visual positioning highly dependent on environmental conditions and low in output frequency, which can lead to inaccurate pose estimation and cumulative errors [26]. In contrast, IMUs (inertial measurement units) offer good dynamic response performance and high pose update frequency but suffer from significant long-term cumulative errors, drift, and poor static characteristics; therefore, to improve the accuracy and applicability range of pose estimation, this paper fuses the pose estimation results of visual odometry with those of the IMU, enhancing the robustness and real-time performance of drone pose estimation.

Visual–IMU fusion methods can be categorized into two types: loose coupling and tight coupling. The basic approach of loose coupling is to process the inertial measurement unit and visual sensors separately, which is relatively simple, requires less computation, and is highly scalable, but the accuracy of the fused pose estimation is relatively poor. The tight coupling approach combines image information with inertial navigation information and calculates the pose information through the fused data. This method is significantly more complex and requires more computation, but can effectively utilize information from each sensor, resulting in the higher accuracy of the fused pose estimation. This paper adopts the tight coupling method in order to achieve the fusion of visual and inertial navigation information.

3.1.1. IMU Pre-Integration

Given the disparity between the capture frequency of cameras and the signal frequency of IMUs, it is essential to consider time synchronization or measurement frequency before fusing IMU data with image information. The method of pre-integration is chosen to overcome the accumulation of errors in the world coordinate system, addressing the issue of the discrepancy in frequency between cameras and IMUs. The fundamental idea behind IMU pre-integration is to extract the invariant items during each optimization iteration

beforehand, thereby reducing the workload of reintegration on each occasion. As depicted in Figure 6, assume that we are in the global coordinate system of the camera, integrating each piece of IMU data from a previous keyframe (at time (t₁)) to a subsequent keyframe (at time (t₂)).

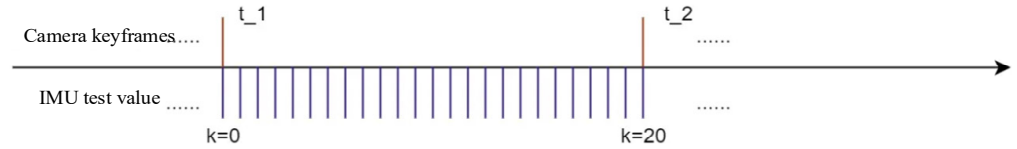


Figure 6. IMU pre-integration schematic diagram.

To optimize the poses and velocities at times (t₁) and (t₂), as well as the IMU bias (denoted as b_a), where the IMU measurement is acceleration, (a_k), the following can be deduced:

$$\begin{aligned} v_k &= v_{k-1} + (a_{k-1} - b_a)\Delta t \\ x_k &= x_{k-1} + v_{k-1}\Delta t + \frac{1}{2}(a_{k-1} - b_a)\Delta t^2 \end{aligned} \tag{3}$$

By integrating directly into the global coordinate system, the following can be inferred:

$$\begin{aligned} v_{t_2} &= v_{t_1} + \sum_{k=0}^{t_2-t_1} (a_k - b_a)\Delta t \\ x_{t_2} &= x_{t_1} + \sum_{k=0}^{t_2-t_1} \left(v_k\Delta t + \frac{1}{2}(a_k - b_a)\Delta t^2 \right) \end{aligned} \tag{4}$$

Since the drone operates in three-dimensional space, the velocity, (v), is directional. Therefore, each time optimization occurs, updating the velocity, (v_{t1}), in the global coordinate system requires recalculating the result within the summation sign.

$$\begin{aligned} \beta &= \sum_k (a_k - b_a)\Delta t \\ \alpha &= \sum_{k=0}^{t_2-t_1} \left[\sum_{l=0}^{k-1} (a_l - b_a)\Delta t \cdot \Delta t + \frac{1}{2}(a_k - b_a)\Delta t^2 \right] \\ &= \sum_{k=0}^{t_2-t_1} \left[\beta_k\Delta t + \frac{1}{2}(a_k - b_a)\Delta t^2 \right] \end{aligned} \tag{5}$$

The calculation is performed as a result of pre-integration. Then, the formula for calculating the position is modified to the following:

$$\begin{aligned} x_{t_2} &= x_{t_1} + \sum_{k=0}^{t_2-t_1} \left(v_k\Delta t + \frac{1}{2}(a_k - b_a)\Delta t^2 \right) \\ &= x_{t_1} + \sum_{k=0}^{t_2-t_1} \left[v_{t_1}\Delta t + \sum_{l=0}^{k-1} (a_l - b_a)\Delta t \cdot \Delta t + \frac{1}{2}(a_k - b_a)\Delta t^2 \right] \end{aligned} \tag{6}$$

Therefore, each time the velocity, v_{t1}, in the global coordinate system is updated, it is only necessary to substitute the results of pre-integration to obtain the new integration results:

$$\begin{aligned} v_{t_2} &= v_{t_1} + \beta \\ x_{t_2} &= x_{t_1} + v_{t_1}(t_2 - t_1) + \alpha \end{aligned} \tag{7}$$

This approach also reduces the necessity of recalculating when the original state information changes, significantly lowering computational complexity. By expanding according to Equation (7) using the Euler integration method, the values of these three variables can ultimately be obtained. This completes the process of IMU value pre-integration. Subsequently, each optimization only requires multiplying the new velocity by time, thus greatly reducing the amount of computation required.

3.1.2. Tight Coupling Optimization Model of Vision Inertial Navigation

As illustrated in Figure 7, visual feature measurements can be obtained through visual sensors. Subsequently, corresponding image feature information is derived using relevant algorithms. By integrating these with IMU measurements from the same time period, more accurate pose and velocity information can be acquired.

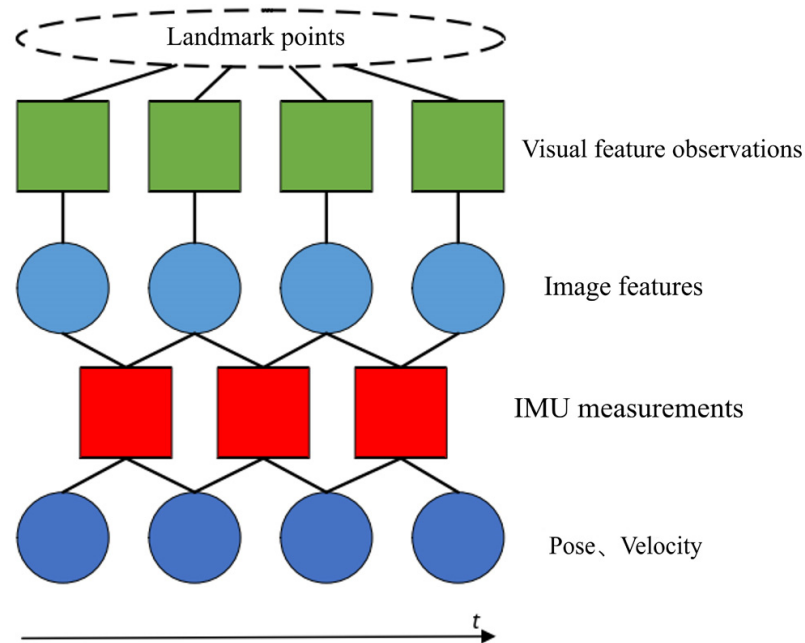


Figure 7. Visual–inertial fusion process.

The specific steps for the visual–inertial fusion are as follows:

1. Based on the IMU motion model, integration can be performed on the IMU for the next image frame, obtaining a predicted value of the IMU increment;
2. The difference between the predicted pre-integration value of the IMU and the observed value obtained from the IMU measurement at the next image frame is calculated, yielding the residual of the inertial data;
3. The residual of the inertial data is introduced into the optimal cost function of visual SLAM, making the optimal cost function a sum of the visual reprojection error and inertial data residual.

Equation (8) represent the motion equation and observation equation that describe the process of the body’s motion. The motion state at time $k + 1$, denoted as $k + 1$, can be predicted based on the state at time k and the IMU measurements between times k and $k + 1$. The observation equation is used to describe the process of the camera detecting visual features at time $k + 1$:

$$\begin{aligned} X_{k+1} &= f(X_k, \mu_{k+1}) + \omega_{k+1} \\ p_{k+1} &= h(p_{k+1}, X_{k+1}) + v_{k+1} \end{aligned} \tag{8}$$

The symbol $X_{k+1} = (R_{k+1}, p_{k+1}, v_{k+1}, b_{k+1}^g, b_{k+1}^a)^T$ represents the system’s state at time $k + 1$, which encompasses the rotation information, velocity information, translation information, and the biases of the gyroscope and accelerometer from time k . The term μ_{k+1} refers to the motion measurement provided by the IMU.

The kinematic process of a body is a cyclic iterative process. Based on the kinematic equations and inputs, it predicts the state at the next moment and uses the sensor’s measurement equations to correct the state at the next moment. In the visual–inertial fusion algorithm, IMU data also act as an input, thereby allowing for more accurate predictions of

the system’s state using the kinematic equations. This process emphasizes the continuous integration of motion data and sensor measurements to refine the system’s understanding of its state over time, leveraging both the predictive power of the motion model and the corrective insights from real-time sensor data.

3.2. GPS Fusion Based on Pose Graph Optimization

This section introduces a sensor fusion algorithm based on pose graph optimization, which is used to integrate the GPS with the visual–inertial positioning algorithm introduced previously. It utilizes an optimization approach to complement the advantages of high local positioning accuracy and global positioning without cumulative errors.

After converting the absolute position data obtained from the GPS sensor into the coordinate system, the position under the east, north, up (ENU) coordinate system, taking the position at the first moment as the origin, can be obtained to determine the drone’s relative position. This is then aligned in time with the relative position data from the visual–inertial odometer obtained through the fusion of visual and IMU sensors, eliminating data with mismatched timestamps to ensure the correspondence of the data (Figure 8).

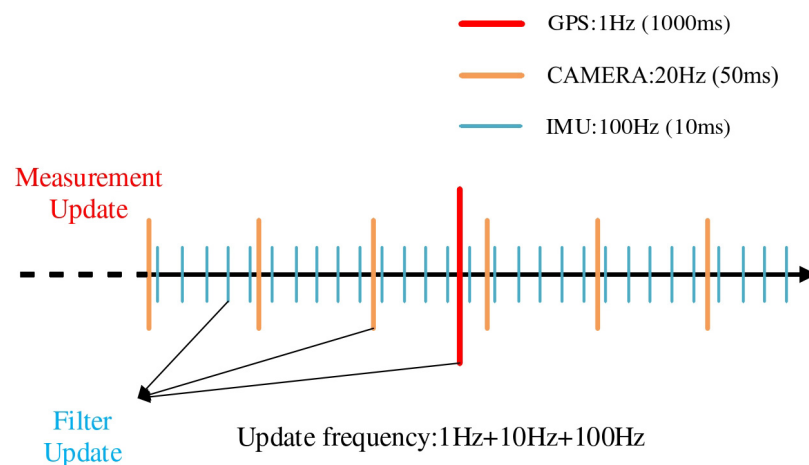


Figure 8. Multisensor data fusion time alignment.

Since the relative poses from the visual–inertial odometer are quite accurate with minimal short-term drift, the covariance of the visual–inertial odometer is set to 0.1. The covariance of the GPS is determined by the strength of the signal. In the fusion algorithm, the weights of the various sensors are dynamically allocated in real-time based on the covariance of GPS and the visual–inertial odometer to achieve better fusion effects [27].

Pose graph optimization is a computational method used in SLAM to optimize the poses of robots. Compared to the commonly used bundle adjustment (BA) optimization in SLAM, pose graph optimization excludes the numerous three-dimensional space points considered in BA optimization, focusing only on the drone’s position and attitude. This significantly reduces the scale of optimization and avoids unnecessary computational power consumption. Figure 9 illustrates the construction of the pose graph.

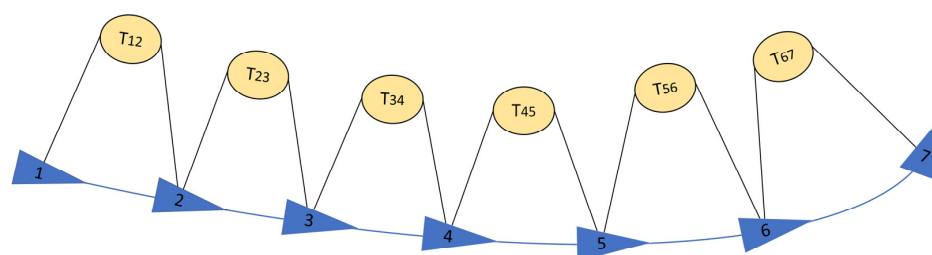


Figure 9. Schematic diagram of the pose graph.

In Figure 9, the triangles represent the drone’s motion posture at each moment, referred to as nodes in the optimization process. The circles represent the change between two consecutive motion postures, known as relative measurements, and in the context of optimization problems they are also called edges. In bundle adjustment (BA) optimization, the number of map points significantly exceeds the number of poses. If the system operates over a long duration, the number of nodes awaiting optimization increases, leading to slower solution speeds. To improve the solution speed, the concept of pose graph optimization emerged. The relative measurements between two motion states can be directly obtained through various sensors carried by the drone (inertial navigation, vision, and GPS).

Assuming that the pose of the UAV at two times is T_i, T_j , and the relative pose change of the UAV between these two times is ΔT_j^i , then there is the following formula:

$$\Delta T_j^i = T_i^{-1}T_j \tag{9}$$

Build an error term:

$$e_{ij} = \Delta T_j^{i-1}T_i^{-1}T_j \tag{10}$$

The objective function of pose graph optimization can be written as follows:

$$F(x) = \min_x \sum_1^n e_{ij}^T \Lambda_{ij} e_{ij} \tag{11}$$

In the above equation, Λ_{ij} represents the information matrix, which is the inverse of the covariance, indicating the weight of each error in the optimization function. If we set the initial point of the optimization problem as \bar{x}_k , and consider an increment added on top of this initial point, then the estimated value between state nodes becomes $F_k(\bar{x}_k + \Delta x)$, while the error value between nodes is $e_k(\bar{x}_k + \Delta x)$. Subsequently, the overall error cost function is linearized for processing:

$$e_k(\bar{x}_k + \Delta x) \approx e_k(\bar{x}_k) + \frac{de_k}{dx_k} \Delta x = e_k + J_k \Delta x \tag{12}$$

For the k -th edge, the optimization function is as follows:

$$\begin{aligned} F_k(\bar{x}_k + \Delta x) &= e_k(\bar{x}_k + \Delta x)^T \Lambda_k e_k(\bar{x}_k + \Delta x) \\ &\approx (e_k + J_k \Delta x)^T \Lambda_k (e_k + J_k \Delta x) \\ &= e_k^T \Lambda_k e_k + 2e_k^T \Lambda_k J_k \Delta x + \Delta x^T J_k^T \Lambda_k J_k \Delta x \\ &= C_k + 2b_k \Delta x + \Delta x^T H_k \Delta x \end{aligned} \tag{13}$$

In the above formula, the constant term is sorted into C_k . In fact, the variation in the final optimization function is as follows:

$$\Delta F_k = 2b_k \Delta x + \Delta x^T H_k \Delta x \tag{14}$$

The core of solving the optimization problem is to find a state variable, Δx , to minimize the variation in the above formula, so the most direct way is to take the derivative of the above formula and then make the derivative of Δx the smallest derivative in the definition domain, and then the optimal solution can be obtained:

$$\frac{dF_k}{d\Delta x} = 2b + 2H_k \Delta x = 0 \Rightarrow H_k \Delta x = -b_k \tag{15}$$

When Δx decreases to a certain extent, the iteration is stopped.

3.3. Digital Twin—Kalman Fusion Algorithm

In multisensor fusion positioning systems, although the addition of sensors can provide more sources of data, enhancing the system's perception of the external environment and positioning accuracy, it also introduces more complexity and uncertainty. From the perspectives of probability theory and systems engineering, as the number of sensors increases the complexity of the system grows linearly or at an even higher rate, leading to a corresponding increase in the probability of failures. Each sensor could potentially become a point of failure, increasing the overall probability of system failure [28]. Moreover, errors between different sensors may also accumulate or amplify each other, affecting the final data fusion quality and positioning accuracy; therefore, this study introduces digital twin technology as an innovative strategy. By constructing a digital twin model of the real drone, it achieves the real-time simulation of the real drone's flight status. This digital twin drone can not only predict and simulate the behavior and state of the real drone but also provide alternative positioning information in case the actual drone encounters sensor failures or data anomalies.

This paper will employ pose optimization techniques based on the extended Kalman filter (EKF) to improve the accuracy of the drone's motion posture. The core of pose optimization techniques based on the EKF lies in the fusion of different information sources. Applying the Kalman filter to pose optimization involves using data generated by the digital twin model as estimated values, combined with real drone sensor data for posture prediction. Through the extended Kalman filter, the position information of the drone being measured in space is continuously corrected, and the filtered position information is fed back to the real drone, thereby enhancing the accuracy and robustness of the drone's motion posture and position information.

In UAV systems, attitude information directly influences the movement direction and velocity vector of a UAV. Given that attitude information is typically of high frequency, accurate attitude estimation can significantly reduce uncertainties in position estimation. This enhancement in accuracy allows a system to exhibit greater robustness against external disturbances and sensor noise; by initially using high-frequency attitude information for attitude estimation, it provides precise direction and orientation references for subsequent low-frequency position estimation.

Take the attitude information of the digital twin UAV at the current time as the following:

$$X = \{ \theta_{\text{roll}}, \theta_{\text{yaw}}, \theta_{\text{pitch}} \} \quad (16)$$

The computational state update equation can be described as a first-order Markov process:

$$\begin{cases} \theta'_{\text{roll}} = -\tau_{\epsilon_e} \theta_{\text{roll}} + \omega_{\epsilon_e} \\ \theta'_{\text{yaw}} = -\tau_{\epsilon_n} \theta_{\text{yaw}} + \omega_{\epsilon_n} \\ \theta'_{\text{pitch}} = -\tau_{\epsilon_\theta} \theta_{\text{pitch}} + \omega_{\epsilon_\theta} \end{cases} \quad (17)$$

where $\theta'_{\text{roll}}, \theta'_{\text{yaw}}, \theta'_{\text{pitch}}$ are the pose state of the target at the next moment; $\tau_{\epsilon_e}, \tau_{\epsilon_n}, \tau_{\epsilon_\theta}$ are the reciprocal of the time constant of the noise process; and $\omega_{\epsilon_e}, \omega_{\epsilon_n}, \omega_{\epsilon_\theta}$ are white noise with zero mean.

The discrete state equation of X from time k to time $k + 1$ is as follows:

$$\begin{cases} X(k+1) = \varphi X(k) + W_X(k) \\ \varphi = \text{diag}(e^{-\tau_{\epsilon_e} T}, e^{-\tau_{\epsilon_n} T}, e^{-\tau_{\epsilon_\theta} T}) \end{cases} \quad (18)$$

where T is the sampling period and $W_X(k)$ is white noise. Assuming that the state noise at each time is uncorrelated, the correlation state covariance matrix can be obtained as follows:

$$\begin{aligned}
 Q &= \text{diag}(q_{11}, q_{12}, q_{13}) \\
 q_{11} &= \sigma_{\epsilon_e}^2 [1 - \exp(-2\epsilon_{\epsilon_e} T)] \\
 q_{12} &= \sigma_{\epsilon_n}^2 [1 - \exp(-2\epsilon_{\epsilon_n} T)] \\
 q_{13} &= \sigma_{\epsilon_\theta}^2 [1 - \exp(-2\epsilon_{\epsilon_\theta} T)]
 \end{aligned}
 \tag{19}$$

The position and attitude state of the previous state point Ω is the observation value:

$$Z = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} P'_e - P_e \\ P'_n - P_n \\ \theta' - \theta \end{pmatrix}
 \tag{20}$$

In this formula, Z represents the three observed values obtained from actual sensor measurements. These values are derived from the sensor's measurement of the position and attitude changes at the current moment. P'_e and P'_n are the observed values in the e-direction and n-direction at the next moment, respectively; P_e and P_n are the observed values in the e-direction and n-direction at the current moment, respectively; θ' is the observed pitch angle at the next moment; and θ is the observed pitch angle at the current moment.

The discrete observation equation is as follows:

$$Z(k) = HX(k) + W_Z(k)
 \tag{21}$$

In the formula, $H = \text{diag}(1, 1, 1)$; $W_Z(k)$ denotes the observation noise. The covariance matrix, R , of the observation noise is represented as follows:

$$R = \begin{bmatrix} \sigma_{em}^2 & \sigma_{enm}^2 & 0 \\ \sigma_{enm}^2 & \sigma_{em}^2 & 0 \\ 0 & 0 & \sigma_\theta^2 \end{bmatrix}
 \tag{22}$$

$$\begin{aligned}
 \sigma_{em}^2 &= \sigma_{xm}^2 \cos^2 \alpha + \sigma_{ym}^2 \sin^2 \alpha \\
 \sigma_{nm}^2 &= \sigma_{xm}^2 \sin^2 \alpha + \sigma_{ym}^2 \cos^2 \alpha \\
 \sigma_{enm}^2 &= (\sigma_{xm}^2 + \sigma_{ym}^2) \cos \alpha \sin \alpha
 \end{aligned}
 \tag{23}$$

After completing the attitude estimation, position estimation is performed. Using the position coordinates and velocity of the digital twin UAV as state variables, the discretized state equation is as follows:

$$X_{k+1} = f[X_k, k] + W_k
 \tag{24}$$

Using the position data obtained from the sensor as measurement information, the observation equation is established as follows:

$$Z_{k+1} = h[X_{k+1}, k + 1] + V_{k+1}
 \tag{25}$$

where W_k and V_{k+1} are uncorrelated noise sequences with a mean of zero, X_{k+1} is the state vector, and Z_{k+1} is the observation vector. By expanding $f[X_k, k]$ around \hat{X}_k using the Taylor series and taking the first-order term, and similarly expanding $h[X_{k+1}, k + 1]$ around $\hat{X}_{k+1|k}$ using the Taylor series and taking the first-order term, we obtain the following:

$$\Phi_{k+1,k} = \left. \frac{\partial f[X_k, k]}{\partial [X_k^T]} \right|_{X_k = \hat{X}_k}
 \tag{26}$$

$$H_{k+1} = \left. \frac{\partial h[X_{k+1}, k+1]}{\partial X_{k+1}^T} \right|_{X_{k+1} = \hat{X}_{k+1|k}} \quad (27)$$

Using the initial position coordinates, velocity, and attitude of the UAV as the initial values of the state vector, and applying the standard Kalman filter, the extended Kalman filter equations are derived as follows, with the filtering process divided into prediction and update stages.

The prior estimate of the state variables at time $k + 1$ is given as follows:

$$\hat{X}_{k+1|k} = f[\hat{X}_k, k] \quad (28)$$

The error covariance at time $k + 1$ is as follows:

$$P_{k+1|k} = \Phi_{k+1} P_k \Phi_k^T + Q_k \quad (29)$$

The Kalman gain is calculated as follows:

$$K_{k+1} = P_{k+1|k} H_{k+1}^T (H_{k+1} P_{k+1|k} H_{k+1}^T + R_{k+1})^{-1} \quad (30)$$

The state estimate is as follows:

$$\hat{X}_{k+1} = \hat{X}_{k+1|k} + K_{k+1} (Z_{k+1} - h[\hat{X}_{k+1|k}, k+1]) \quad (31)$$

The error covariance update at time $k + 1$ is as follows:

$$P_{k+1} = (I - K_{k+1} H_{k+1}) P_{k+1|k} (I - K_{k+1} H_{k+1})^T + K_{k+1} R_{k+1} K_{k+1}^T \quad (32)$$

4. Experiments and Comparative Analysis

4.1. Introduction to the Experimental Platform

For data collection and conducting real flight experiments, a quadcopter drone hardware platform was initially established, as illustrated in Figure 10. This platform was equipped with various sensors as previously described, including an inertial measurement unit (IMU), GPS, and depth cameras. Simultaneously, the sensor models and measurement frequencies used in the digital twin simulation platform matched those of the actual sensors. The GPS provided coarse positioning information at a frequency of 1 Hz, the IMU supplied high-speed dynamic attitude change data at 100 Hz, and the visual sensors delivered visual-assisted positioning information at a rate of 20 Hz.

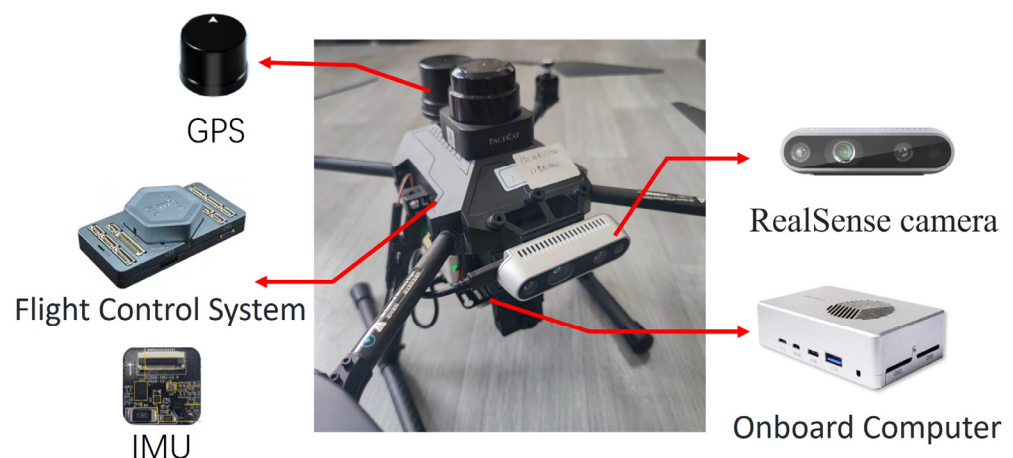


Figure 10. UAV hardware system.

Furthermore, to validate the effectiveness of the digital twin system, this study first established a virtual external testing environment as a primary prerequisite for digital twin testing. To conduct more accurate experiments with the drone digital twin system, the actual experiment environment the school playground was modeled in multiple dimensions. Three-dimensional scene real images and model diagrams are shown in Figure 11.

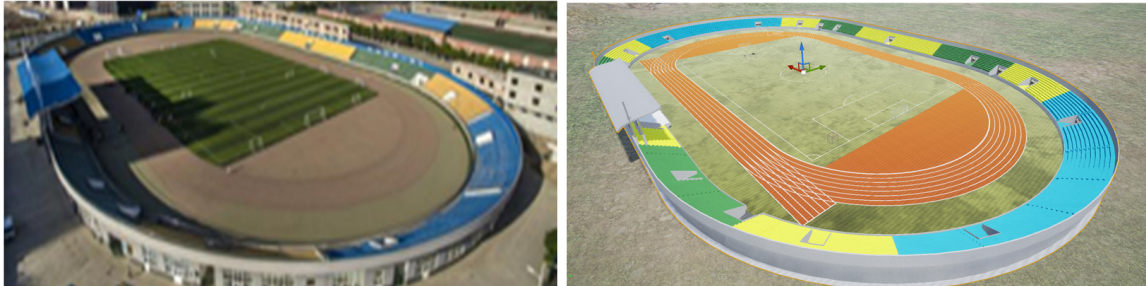


Figure 11. Realistic image (left). Twin scene image (right).

The experimental method employs a multisource multimodal data fusion architecture based on digital twin technology, involving two main components: a physical drone and a digital twin platform, as illustrated in Figure 12. On the physical drone end, various sensors such as GPS, IMU, and RealSense cameras collect pose data in real time. These data are transmitted to the drone’s onboard computer for processing and multimodal fusion. The ground-based digital twin platform serves as the virtual part of the experiment, receiving the drone’s location information in real time, using it as a reference value, and combining it with the digital twin model to calculate the drone’s pose. The calculated pose information, in the form of estimated values, is transmitted back to the drone via data transmission modules using UDP network and MAVLink communication protocols. The drone’s onboard computer receives these estimated values and further fuses them with the measurements obtained from its own multisensor fusion using the extended Kalman filter (EKF) algorithm to optimize the accuracy of pose estimation. The extended Kalman filter, through its iterative prediction and updating process, corrects the drone’s real-time pose, thereby reducing estimation errors.

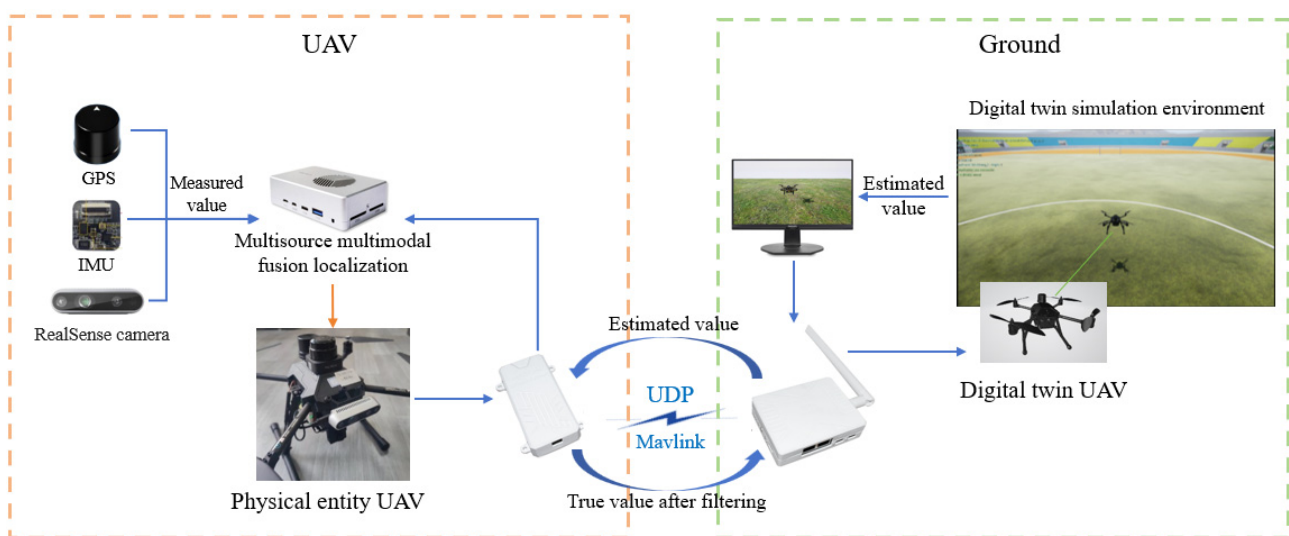


Figure 12. Data fusion experiment frame diagram.

The entire experimental process forms a closed-loop system, where the real-time interaction between the physical drone and the digital twin platform ensures the continuous

synchronization and precise calibration of pose data. This method, by fusing physical multisensor data with virtual data from the digital twin platform, aims to enhance the accuracy of pose estimation and validate the effectiveness as well as the reliability of multisource, multimodal data in practical applications.

4.2. Comparison of Integrated Navigation Experiments

The digital twin environment for this experiment utilizes UE4 to simulate a real campus playground environment. While the real drone is conducting actual flight operations, the digital twin drone runs synchronously in the twin environment, collecting sensor data. The frequencies of the sensors used in the digital twin simulation match those of the real sensors: the inertial sensor sampling rate is 100 Hz, the GPS sensor frequency is 1 Hz, and the vision sensor frequency is 20 Hz. This paper focuses on a comparative analysis of the integrated application of various combined navigation methods. It compares three different combination navigation strategies: visual odometry/inertial measurement unit (VO/IMU), global positioning system/inertial measurement unit (GPS/IMU), and global positioning system/visual odometry/inertial measurement unit (GPS/VO/IMU). By analyzing their performance in positioning accuracy, theoretical foundations can be provided for the selection of UAV navigation systems.

The experimental results in Figure 13 show that the GPS/VO/IMU combination navigation method exhibits smaller positioning errors and higher stability. This finding suggests that the fusion of GPS, VO, and IMU data in drone positioning can effectively suppress errors that may arise from individual systems, reduce positioning fluctuations caused by the performance variability of a single system, and significantly enhance overall navigation accuracy. Compared to using VO/IMU or GPS/IMU alone, the GPS/VO/IMU combination not only provides more accurate location information but also maintains stable system performance under complex environmental conditions.

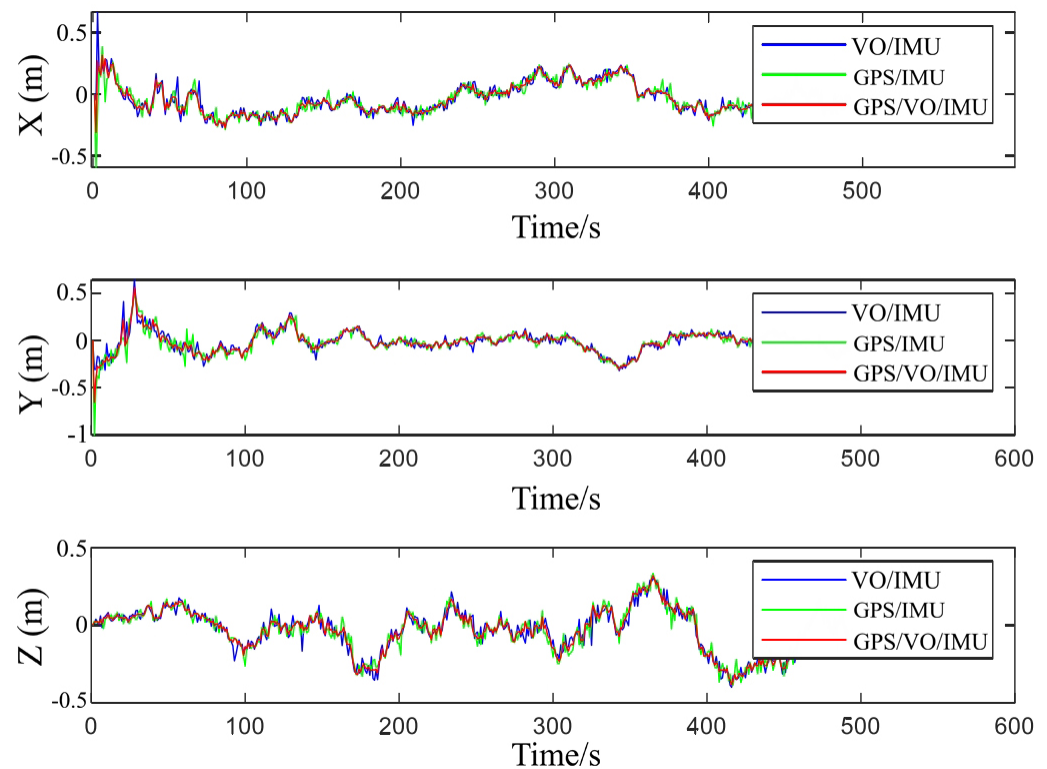


Figure 13. The error curves of integrated navigation in the x, y, and z directions. The blue line represents the navigation error of the visual and IMU fusion, the green line represents the navigation error of the GPS and IMU fusion, and the red line represents the navigation error of the GPS, IMU, and visual fusion.

These experimental results further confirm the importance of multimodal information fusion in improving the robustness of positioning systems. Especially when GPS signals are limited or the accuracy of the VO system is affected by indistinct environmental features, multisource information fusion ensures the continuity and accuracy of the system. Additionally, the GPS/VO/IMU combination navigation, by complementing the strengths and weaknesses of each system, not only enhances positioning accuracy but also improves the system's adaptability to dynamic changes and resistance to interference, providing strong support for the stable operation of drones in variable environments.

4.3. Analysis of Navigation Stability Experiment

Figure 14 showcases the distribution of positional and attitudinal errors and the covariance of the drone system before applying the data fusion technique proposed in this paper. From the figure, it can be observed that although the system can maintain errors within a relatively small range in the initial phase, indicating a certain degree of accuracy and stability in the short term, the errors in position and attitude exhibit a trend of gradual increase over time, manifesting as a systematic drift phenomenon. This drift leads to the divergence of covariance, reflecting an increase in the uncertainty of system state estimation. The performance degradation of the system indicates a lack of effective error correction mechanisms during long-term operation, ultimately leading to the inability of the drone's navigation system to maintain a normal working state, thereby validating the insufficiency of single-sensor or unoptimized multisensor systems in maintaining long-term positioning stability.

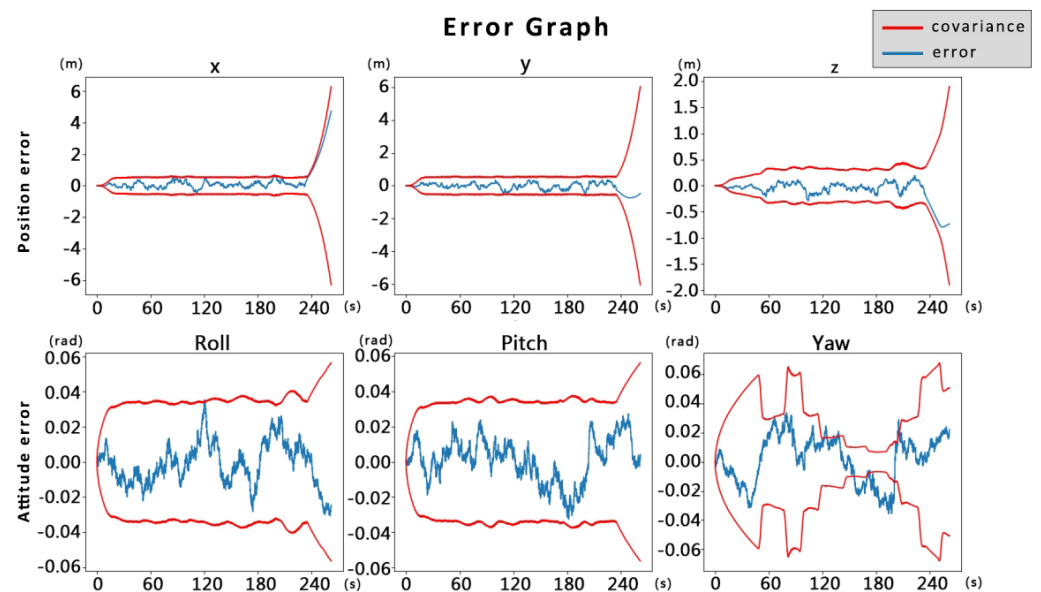


Figure 14. Visual positioning error and covariance line chart. The red line represents the covariance curve, and the blue line represents the error line.

In contrast, Figure 15 shows the error distribution and covariance after implementing the data fusion method proposed in this study. Compared to Figure 14, the covariance matrix maintained smaller fluctuations throughout the experiment period without significant divergence. This proves that the multimodal fusion positioning method proposed in this paper can effectively solve the problem of system state divergence caused by visual positioning errors, increasing the stability and robustness of state estimation, and improving accuracy based on actual operational results. This means that, under diverse operational conditions and various disturbances that may be encountered, the drone's navigation system can still maintain high positioning accuracy, ensuring the reliability and safety of the drone's mission execution.

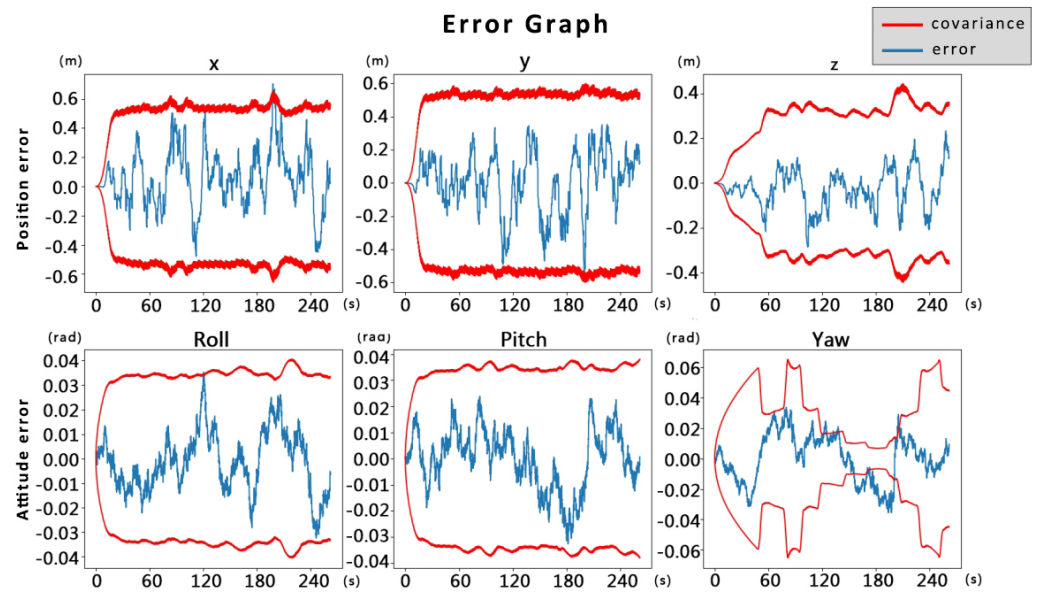


Figure 15. Multimodal data fusion navigation error and covariance line. The red line represents the covariance curve, and the blue line represents the error line.

4.4. Experimental Analysis of Fault-Tolerant Navigation

It is evident from Figure 16 that a gradual fault was applied between 50 s and 150 s, while a sudden fault was applied between 250 s and 350 s. The component applied pertains to the drone’s x-axis direction, with little difference observed in the y-axis and z-axis components; however, the occurrence of the fault and the distinction between the two can be clearly seen in the x-axis component.

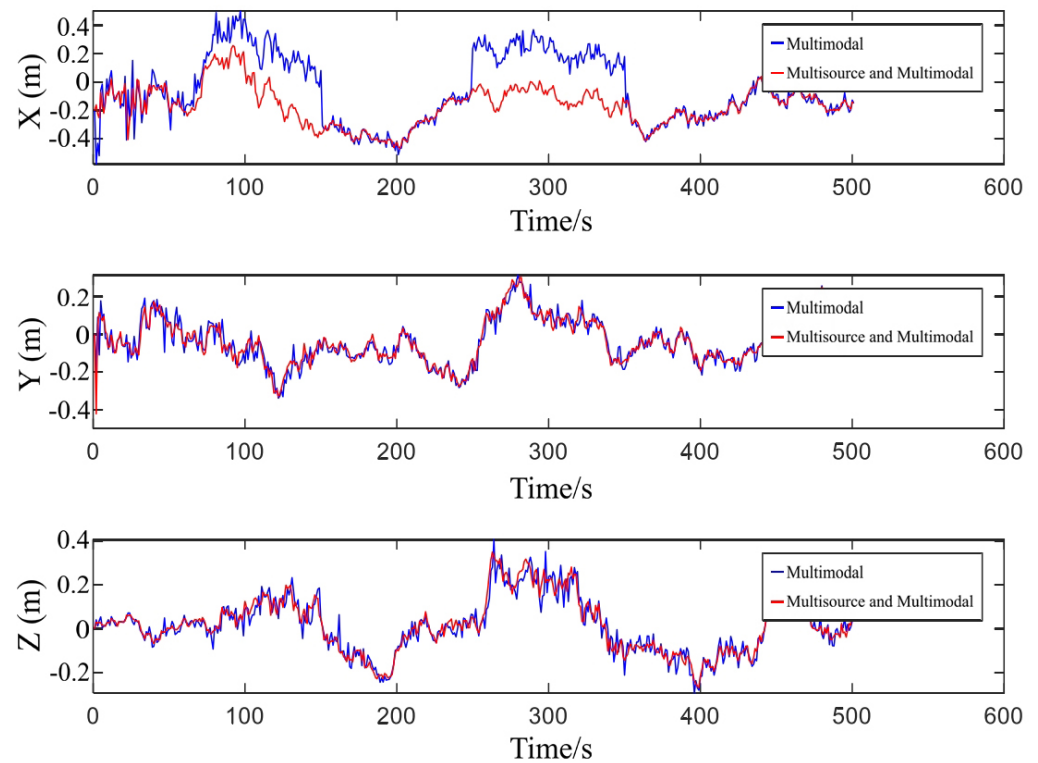


Figure 16. The error line graph comparing multimodal with multisource, multimodal (incorporating digital twin drone sources).

In the absence of a digital twin drone data source and the utilization of a multisource, multimodal navigation algorithm, the drone system experiences abrupt changes, especially in the x-axis direction, when subjected to the applied faults, leading to system instability. This instability manifests as rapid changes in position and attitude, thereby affecting the normal operation of the drone; however, the drone system, incorporating the multisource multimodal navigation algorithm, demonstrates significant stability. In the face of gradual faults, despite potential delays in detection that cause some variation in the response of the multisource multimodal navigation algorithm, the variation is minimal, and the system can quickly return to a stable state. In the case of sudden faults, the multisource, multimodal navigation algorithm exhibits high robustness, with almost no impact, allowing for a smooth transition and ensuring the stability of the combined navigation system.

PX4 is an open-source flight control system that is widely used in drones and other unmanned systems. Its navigation module employs integrated navigation methods, which are more complex than the standalone GPS/IMU algorithm as it fuses data from multiple sensors. This paper uses PX4 as a benchmark with which to validate the effectiveness of the designed fusion algorithm. We conducted waypoint mission flights near tall buildings, and the PX4 flight control data were transmitted to the ground station in real-time via a wireless link to obtain the trajectory. The experimental results are shown in Figure 16. In Figure 17b, the area marked with an orange ellipse indicates the trajectory when the drone was near tall buildings and the number of satellites was fewer than six. It can be seen that the position data from the PX4 flight control system, which relies on GPS signals, began to drift and fluctuate, whereas the position estimated by the fusion algorithm only exhibited minor oscillations and remained stable. Additionally, in Figure 17a, the three-dimensional coordinates represent the relative x, y, and z positions of the drone. Overall, compared to the PX4 flight control trajectory, the drone using the data fusion algorithm maintained a smooth and consistent trajectory throughout the flight, reducing sharp fluctuations and abrupt changes. This indicates that the data fusion technology effectively filters out noise and errors in sensor data.

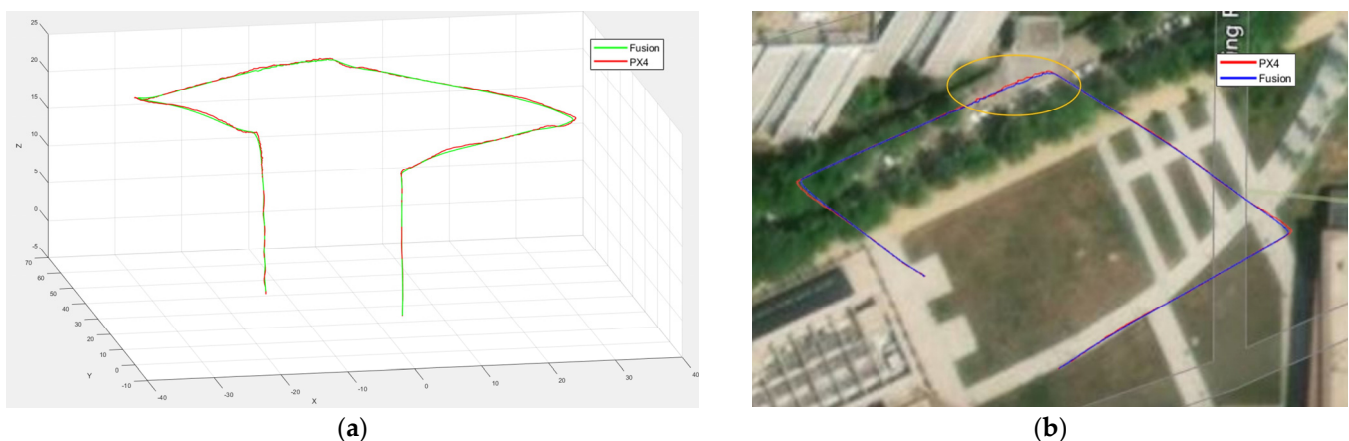


Figure 17. Actual flight trajectories of the drone near tall buildings. (a) Comparison of the three-dimensional flight trajectory lines of the drone. The red dashed line represents the trajectory data from the PX4 flight control system, while the green line represents the trajectory data from the fusion-based system. (b) Comparison of the two-dimensional planar trajectories on the ground station. The blue line represents the trajectory data from the fusion-based system, and the red line represents the flight trajectory data provided by the PX4 flight control system, the area marked with an orange ellipse indicates the trajectory when the drone was near tall buildings and the number of satellites was fewer than six.

Specifically, the digital twin technology provides a robust reference for data fusion while simulating the real drone; by utilizing predictive data generated from the digital twin model and integrating it with real sensor data, the algorithm effectively combats the

instability caused by unexpected faults. In facing sensor errors or external disturbances, the system can adjust state estimates more stably. Moreover, this method also enhances positioning accuracy. As shown in Figures 16 and 17, the fluctuation range of positional errors is relatively smaller and more concentrated, indicating the precision advantage of the fusion method.

5. Conclusions and Future Work

This paper first discussed the challenges of maintaining positioning accuracy and stability during the continuous maneuvers of small drones in complex environments. To address this, it proposed a fusion positioning method that incorporates three different sensor technologies: GPS, inertial navigation, and visual navigation. After collecting relevant data from each sensor, a tightly coupled integration optimized the fusion of visual and inertial data, which were then further integrated with GPS data based on pose graph optimization principles. Several experimental comparisons of navigation combinations and navigational stability analyses were carried out. The experiments proved that the combined navigation system utilizing GPS/VO/IMU enhances positional accuracy by leveraging the strengths and mitigating the weaknesses of each individual system. This combination not only provides more precise location information but also maintains a stable system performance in complex environmental conditions, resulting in a smoother UAV flight trajectory; however, as the number of sensors increases, the system's complexity grows linearly or even at a higher rate, which in turn increases the likelihood of system failures. Consequently, this research introduced a digital twin UAV data source. By using the data generated by the digital twin model as an estimation input, combined with real UAV sensor data via the extended Kalman filter (EKF) algorithm, the approach corrects for positional information deviations caused by sensor noise and environmental disturbances. The filtered positional information is then fed back into the control system of the real UAV to achieve real-time corrections of the UAV's movement attitude and positional information. The method has been empirically proven to significantly improve the UAV's positioning accuracy, exhibiting a high level of robustness and ensuring the stability of the combined navigation system as well as the reliability of the navigation data.

In conclusion, the data fusion method proposed in this study has been theoretically and experimentally proven to be effective in addressing UAV positioning issues in complex dynamic environments. The main contributions of this paper include the following:

- (i) Proposing a multimodal data fusion method: This paper introduces a positioning method that tightly couples and optimally fuses data from GPS, IMU, and visual sensors. This approach successfully addresses the problem of insufficient accuracy from single sensors, significantly enhancing UAV positioning accuracy and stability, especially in complex and dynamic environments.
- (ii) Introducing digital twin technology: To address the increased probability of failure due to the rising number of sensors, this study establishes a high-fidelity UAV model within the digital twin. By introducing the digital twin UAV as a robust reference in the data fusion process, this innovation effectively reduces the impact of abrupt changes in sensor data on positioning accuracy. It also provides a solid foundation for the long-term stable operation of the system and offers important references as well as research directions for the design and optimization of future UAV navigation systems.
- (iii) Application of the extended Kalman filter algorithm: This study employs the extended Kalman filter algorithm to achieve efficient multisource data fusion. The extended Kalman filter algorithm can handle state estimation problems in nonlinear systems, correcting position information deviations caused by sensor noise and environmental interference in real time. This ensures the stable flight and precise navigation of UAVs in various complex scenarios, significantly improving positioning accuracy and system robustness in challenging environments.

Future inquiries shall pivot towards an enhanced exploration of fusion algorithm optimization tailored to accommodating more intricate environments and rigorous application

demands, harnessing cutting-edge computational technologies such as edge computing or cloud computing to expedite data processing, diminishing system response latencies, and bolstering real-time application performance. An in-depth examination and refinement of the extant Kalman filter algorithms are envisaged to fortify their adaptability and precision in the management of data within highly dynamic and convoluted environmental contexts. Moreover, a deep dive into the application of digital twin technology within UAV locational matrices, particularly within the realms of fault diagnostics and prognostic maintenance, is anticipated. The construction of digital twin models for UAVs will permit the real-time surveillance of UAV operational statuses, the prognostication of potential malfunctions, and the pre-emptive enactment of preventative measures, thereby augmenting system stability and reliability.

Author Contributions: Conceptualization, S.Q. and Y.F.; methodology, J.C.; software, J.C.; validation, S.Q.; formal analysis, Y.Q.; investigation, Z.C.; resources, Y.F. and Z.C.; data curation, Y.Q. and X.M.; writing—original draft preparation, J.C.; writing—review and editing, S.Q.; visualization, X.M.; supervision, S.Q.; project administration, Y.F.; funding acquisition, Y.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the National Foreign Expert Program of the Ministry of Science and Technology (grant no. G2023041037L) and the Special Research Program of the Shaanxi Provincial Department of Education (grant no. 19JK0414).

Data Availability Statement: Data are contained within the article.

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

References

1. Xiao, H. *Research and Implementation of UAV Target Detection System for LINE patrol Environment Now*; Beijing University of Technology: Beijing, China, 2021. [\[CrossRef\]](#)
2. Peng, Z. *Research and Implementation of Indoor Autonomous Navigation System for Four-Rotor UAV Based on Vision*. Master's Thesis, University of Electronic Science and Technology of China, Chengdu, China, 2019.
3. Yan, G.; Weng, J. *Strapdown Inertial Navigation Algorithm and Integrated Navigation Principle*; Northwestern Polytechnical University Press: Xi'an, China, 2019; pp. 127–128.
4. Wang, T.; Cai, Z.; Wang, Y. Indoor vision/inertial navigation method for UAV. *J. Beijing Univ. Aeronaut. Astronautics* **2018**, *44*, 176–186.
5. Usenko, V.; Von Stumberg, L.; Pangercic, A.; Cremers, D. Real-time trajectory replanning for MAVs using uniform B-splines and a 3D circular buffer. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, ON, Canada, 24–28 September 2017; pp. 215–222.
6. Alzugaray, I.; Teixeira, L.; Chli, M. Short-term UAV path-planning with monocular-inertial SLAM in the loop. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 2739–2746.
7. Ruochen, B. *Research on Multi-Source Heterogeneous Adaptive Integrated Navigation Technology in Complex Environment*. Master's Thesis, Shandong University, Jinan, China, 2020. [\[CrossRef\]](#)
8. Jing, P. *Research on UAV Localization Technology Based on Multi-Perception Information Fusion*. Master's Thesis, South China University of Technology, Hangzhou, China, 2020. [\[CrossRef\]](#)
9. Jin, X.B.; Sun, S.; Wei, H.; Yang, F. Advances in Multi-Sensor Information Fusion: Theory and Applications 2017. *Sensors* **2018**, *18*, 1162. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Wang, R.A.; Chen, Y.B.; Li, Y.C.; Xu, P.; Peng, S. High-precision initialization and acceleration of particle filter convergence to improve the accuracy and stability of terrain aided navigation -ScienceDirect. *ISA Trans.* **2020**, *110*, 172–197.
11. Tang, L.; Tang, X.; Li, B.; Liu, X. A review of fusion algorithms for multi-source fusion navigation systems. *Glob. Position. Syst.* **2018**, *43*, 39–44.
12. Tang, C. Environmental Modeling and Exploration of small UAV based on RGB-D Data in Indoor Environment. *Soochow Univ.* **2020**. [\[CrossRef\]](#)
13. Ye, D.; Qilin, F.; Jian, Z. Discussion on digital twin battlefield construction. *Prot. Eng.* **2020**, *42*, 58–64.
14. Wang, P.; Yang, M.; Zhu, J.; Peng, Y.; Li, G. Digital Twin-Enabled Online Battlefield Learning with Random Finite Sets. *Comput. Intell. Neurosci.* **2021**, *2021*, 5582241. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Du, H.; Wang, W.; Xu, C.; Xiao, R.; Sun, C. Real-Time Onboard 3D State Estimation of an Unmanned Aerial Vehicle in Multi-Environments Using Multi-Sensor Data Fusion. *Sensors* **2020**, *20*, 919. [\[CrossRef\]](#)
16. Causa, F.; Fasano, G. Improved In-Flight Estimation of Inertial Biases through CDGNSS/Vision Based Cooperative Navigation. *Sensors* **2021**, *21*, 3438. [\[CrossRef\]](#)

17. Luo, H.; Zhang, J.; Ning, Y.; You, Y. Application requirements of digital twins in submarine combat system. *Digit. Ocean. Underw. Attack Def.* **2021**, *4*, 233–237.
18. Wang, Q.; Cui, X.; Li, Y.; Ye, F. Performance Enhancement of a USV INS/CNS/DVL Integration Navigation System Based on an Adaptive Information Sharing Factor Federated Filter. *Sensors* **2017**, *17*, 239. [[CrossRef](#)] [[PubMed](#)]
19. Wang, Q.; Guo, Z.; Zhang, M.; Cui, X.; Wu, H.; Jia, L. Research on pedestrian location based on dual MIMU/magnetometer/ultrasonic module. In Proceedings of the Position, Location and Navigation Symposium (PLANS), Monterey, CA, USA, 23–26 April 2018; pp. 565–570.
20. Yan, J.; He, G.; Basiri, A.; Hancock, C. Vision-aided indoor pedestrian dead reckoning. In Proceedings of the 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Houston, TX, USA, 22–25 May 2018; pp. 1–6.
21. Soliman, A.; Al-Ali, A.; Mohamed, A.; Gedawy, H.; Izham, D.; Bahri, M.; Erbad, A.; Guizani, M. AI-based UAV navigation framework with digital twin technology for mobile target visitation. *Eng. Appl. Artif. Intell.* **2023**, *123 Pt B*, 106318. [[CrossRef](#)]
22. Yang, Y.; Liu, X.; Zhang, W.; Liu, X.; Guo, Y. A Nonlinear Double Model for Multisensor-Integrated Navigation Using the Federated EKF Algorithm for Small UAVs. *Sensors* **2020**, *20*, 2974. [[CrossRef](#)]
23. Lee, J.-C.; Chen, C.-C.; Shen, C.-T.; Lai, Y.-C. Landmark-Based Scale Estimation and Correction of Visual Inertial Odometry for VTOL UAVs in a GPS-Denied Environment. *Sensors* **2022**, *22*, 9654. [[CrossRef](#)]
24. Strohmeier, M.; Montenegro, S. Coupled GPS/MEMS IMU Attitude Determination of Small UAVs with COTS. *Electronics* **2017**, *6*, 15. [[CrossRef](#)]
25. George, A.; Koivumäki, N.; Hakala, T.; Suomalainen, J.; Honkavaara, E. Visual-Inertial Odometry Using High Flying Altitude Drone Datasets. *Drones* **2023**, *7*, 36. [[CrossRef](#)]
26. Luo, Y.; Li, J.; Yu, C.; Xu, B.; Li, Y.; Hsu, L.-T.; El-Sheimy, N. Research on Time-Related Errors Using Allan Variance in a Kalman Filter Applicable to Vector-Tracking-Based GNSS Software-Defined Receiver for Autonomous Ground Vehicle Navigation. *Remote Sens.* **2019**, *11*, 1026. [[CrossRef](#)]
27. Zhuang, Y.; Yang, J.; Qi, L.; Li, Y.; Cao, Y.; El-Sheimy, N. A pervasive integration platform of low-cost MEMS sensors and wireless signals for indoor localization. *IEEE Internet Things J.* **2018**, *5*, 4616–4631. [[CrossRef](#)]
28. Al-Sharman, M.K.; Emran, B.J.; Jaradat, M.A.; Najjaran, H.; Al-Husari, R.; Zweiri, Y. Precision landing using an adaptive fuzzy multi-sensor data fusion architecture. *Appl. Soft Comput.* **2018**, *69*, 149–164. [[CrossRef](#)]

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.