Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 14, No. 3, 2025

Research on Improved Indoor Localization Algorithm for UAVs Based on VINS

Xi Zhang[1, *], Nianqing Tan[2]

[1] School of Avionics and Electrical Engineering, Civil Aviation Flight University of China, Guanghan, Sichuan 618307, China
[2] School of Avionics and Electrical Engineering, Civil Aviation Flight University of China, Guanghan, Sichuan 618307, China
[*] Corresponding author: Xi Zhang (Email: 1815065038@qq.com)

Abstract: To address the limitations of UAV localization accuracy in complex indoor environments, an improved vision-inertial fusion-based UAV indoor positioning method was proposed. The conventional Vision-aided Inertial Navigation System (VINS)-Fusion algorithm was constrained by performance limitations under weak texture and dynamic lighting conditions, leading to degraded system positioning accuracy. To enhance positioning performance, Shi-Tomasi feature points were replaced with ORB (Oriented FAST and Rotated BRIEF) feature points. The binary BRIEF descriptor was employed to improve feature distinctiveness, and the gray centroid method was applied to assign orientation information, ensuring rotational invariance and enhancing image matching robustness in high-speed motion scenarios. Furthermore, a deep learning-driven Feature Booster technique was innovatively introduced to optimize feature descriptors. Through self-enhancement and interactive enhancement mechanisms, the correlations among features were fully exploited, effectively improving descriptor distinctiveness. Finally, a UAV hardware platform based on the RK3588 chip was developed, and a motion capture system was utilized to fine-tune UAV attitude control parameters. Simulations using the EuRoc dataset and real-world indoor flight experiments were conducted to verify the positioning performance and feasibility of the improved algorithm. Experimental results demonstrated that, in complex indoor environments, the root mean square error (RMSE) of the improved algorithm was reduced by 12.24% compared to the original VINS-Fusion algorithm, achieving higher positioning accuracy under dynamic lighting and sparse texture conditions.

Keywords: Feature boosting, uav indoor localization, visual-inertial odometry, vision-inertial fusion.

1. Introduction

With the rapid industrialization of the low-altitude economy [1], UAV technology is increasingly being applied in various fields such as logistics and transportation [2], infrastructure inspection [3], and disaster assessment [4]. However, achieving accurate UAV localization in indoor environments or signal-constrained scenarios remains a significant challenge [5]. Such unstructured environments not only suffer from the inherent deficiency of satellite navigation signal loss but also introduce complex disturbances such as sparse visual features and unstable lighting conditions, making traditional localization algorithms insufficient for navigation requirements. Therefore, achieving high-precision UAV localization under GNSS-denied conditions has become a critical research focus.

Current indoor localization techniques include LiDAR-based [6] and ultra-wideband (UWB)-based [7] methods. Although these approaches provide high accuracy, they require additional sensing equipment or pre-deployed base stations, increasing system deployment costs. In contrast, vision-inertial fusion-based localization methods [8] have attracted considerable attention due to their compact sensor size and ease of integration. Visual sensors provide rich environmental feature information, while an inertial measurement unit (IMU), leveraging its high-frequency data acquisition capability, enhances short-term localization accuracy and system real-time performance [9]. The complementary fusion of these two modalities effectively mitigates the limitations of standalone sensors, reduces accumulated errors, and improves localization capability during high-speed UAV motion [10].

VINS-Fusion is a classic vision-inertial fusion algorithm [11] that achieves efficient pose estimation through frontend feature extraction and tracking, coupled with backend sliding window optimization [12]. However, its performance is highly dependent on the accuracy of frontend feature point extraction and the robustness of feature matching. VINS-Fusion employs the Shi-Tomasi algorithm for feature point detection, which is prone to detection deficiencies and cumulative matching errors in weak-texture and dynamic lighting environments. These limitations affect the accuracy of sliding window optimization, thereby reducing overall system localization performance.

To address these challenges, this study proposes a Feature-Boosted Vision-Inertial Navigation System (FB-VINS) designed for UAV localization in complex indoor environments. The proposed method improves upon the Shi-Tomasi feature point detection module in VINS-Fusion by introducing feature-aware algorithmic optimizations in the visual frontend. A deep learning-based Feature Booster framework [13] is employed to construct a feature enhancement mechanism, dynamically amplifying texture features and their descriptors in key regions through multi-dimensional feature response analysis, without significantly increasing computational complexity.

Furthermore, a UAV hardware platform based on the RK3588 chip was developed, featuring high computational power and low power consumption, ensuring real-time execution of the improved algorithm. Simulations using the EuRoc dataset, along with real-world indoor flight experiments, were conducted to systematically evaluate the localization performance of the proposed method. Experimental results demonstrate that FB-VINS outperforms the conventional VINS-Fusion algorithm in complex indoor environments.

2. Materials and Methods

2.1. Algorithm Implementation Architecture

The proposed algorithm architecture consists of three major modules: visual-inertial frontend, nonlinear optimization backend, and loop closure detection. These components work collaboratively to achieve high-precision and robust UAV indoor localization. The system framework is illustrated in Figure 1.

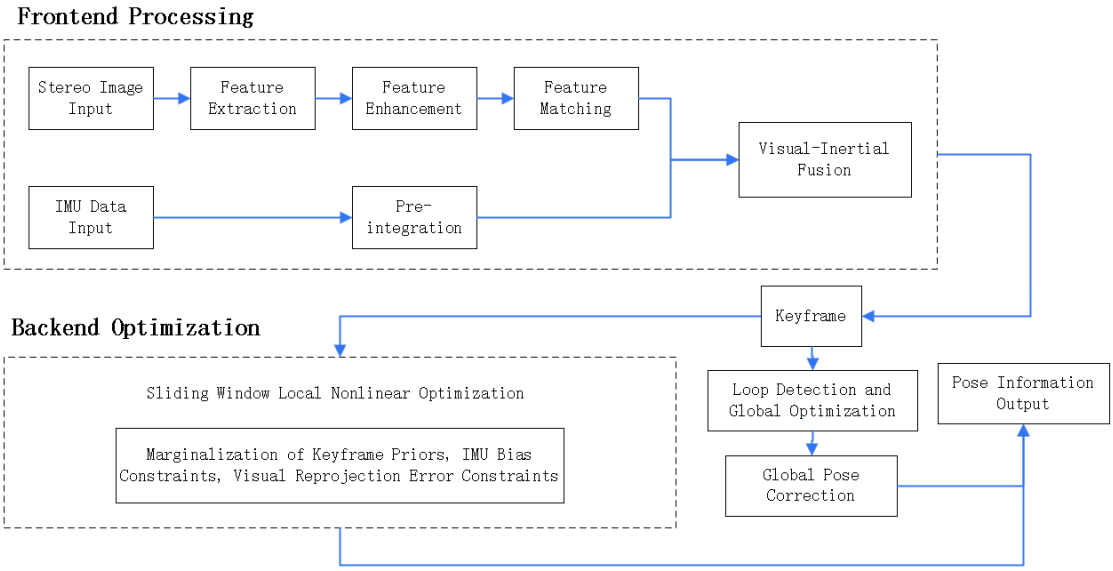

Figure 1. FB-VINS System Framework

2.1.1. Visual-Inertial Frontend

The visual frontend is responsible for extracting high-quality feature points and their descriptors from images and providing the matching results along with depth information to the backend.

During the feature extraction stage, the ORB (Oriented FAST and Rotated BRIEF) algorithm is employed for feature processing [14]. ORB is based on FAST corner detection and incorporates orientation information, generating 256-bit binary descriptors with rotational invariance, ensuring both efficiency and robustness [15]. To address the degradation of ORB feature extraction in weak-texture environments, a deep learning-driven Feature Booster technique is innovatively introduced. This method selectively enhances feature descriptors while maintaining real-time performance, further improving feature quality.

In the feature matching stage, an initial filtering process based on Hamming distance is applied, followed by RANSAC (Random Sample Consensus) with epipolar geometry constraints to eliminate incorrect matches [16].

Finally, a tightly coupled visual-inertial odometry (VIO) framework is adopted, where IMU preintegration data is fused with visual observations to achieve preliminary pose estimation, providing reliable constraints for backend optimization.

2.1.2. Nonlinear Optimization Backend

The nonlinear optimization backend adopts a keyframe-based sliding window tightly coupled framework to jointly optimize the UAV's pose and environment map. The system maintains a fixed-size sliding window, which includes a certain number of keyframes. Each keyframe integrates pose parameters, map point coordinates, and covariance information to ensure consistency and accuracy in localization [17].

The optimization error model consists of three components: visual reprojection error, IMU preintegration error, and prior constraint error [18]. The visual reprojection error is formulated based on the enhanced ORB feature matching results provided by the improved visual frontend. It projects the observed 3D points onto the image plane and compares them with actual observations. This step constrains both the pose of keyframes and the positions of feature points, ensuring accurate feature localization.The IMU preintegration error is constructed using acceleration and angular velocity data between consecutive keyframes. This motion constraint provides high-frequency pose estimation, which contributes to improved localization accuracy and robustness in challenging environments.

As new keyframes are continuously added, the sliding window dynamically updates through a marginalization strategy, which removes outdated keyframes with information entropy below a predefined threshold. This approach effectively enhances computational efficiency and system robustness, enabling stable UAV localization in complex indoor environments.

2.1.3. Loop Closure Detection

Loop closure detection is employed to correct accumulated errors and enhance global consistency [19]. Its primary task is to detect loop closures between the current keyframe and historical keyframes, incorporating loop constraints to optimize the global trajectory.

When a new keyframe is introduced into the sliding window, the loop closure detection module utilizes the Bag-of-Words (BoW) model to match image features between keyframes [20], assessing potential loop closure relationships between the current and historical keyframes.

Once a loop closure is confirmed, the system calculates the relative pose between the current keyframe and the loop

keyframe and integrates this as a loop constraint into the global pose optimization. Through global optimization, the system adjusts keyframe poses to maintain global consistency while preserving the smoothness of local trajectories. This process effectively suppresses pose drift caused by long-term operations, ultimately improving the overall trajectory localization accuracy.

2.2. Improved Frontend Design

To address the decline in localization accuracy of the VINS-Fusion visual frontend in low-light and weak-texture environments, a series of improvements have been made. These enhancements focus on improving the quality of feature descriptors and increasing matching confidence, thereby strengthening the UAV's localization capability in complex environments.

2.2.1. ORB Feature Point Detection

The traditional VINS-Fusion system utilizes the Shi-Tomasi corner detection algorithm in its visual frontend, which performs reliably in texture-rich environments. However, its feature detection rate significantly declines in weak-texture scenarios or under dynamic lighting conditions. To address this issue, the ORB (Oriented FAST and Rotated BRIEF) feature detection algorithm is introduced, which simultaneously optimizes feature extraction and descriptor construction, providing high-quality observational input for backend sliding window optimization.

ORB feature detection is based on the FAST (Features from Accelerated Segment Test) corner detection method, which identifies feature points by analyzing intensity variations within a circular neighborhood. For a given image I and any pixel p with an intensity value of $I(p)$, the intensity threshold for the circular neighborhood $N(p)$ is denoted as δ. If at least N consecutive pixels within the neighborhood satisfy the following condition, p is considered a corner:

$$\left|I(p)-I(p_i)\right|>\delta, p_i \in N(p) \tag{1}$$

To improve feature quality, the Harris corner scoring mechanism is introduced. By analyzing the gradient distribution within the pixel neighborhood, a self-correlation matrix is constructed, and its determinant and trace are combined as a selection criterion. This prioritizes feature points with significant gradient variations, calculated as follows:

$$R=\det(M)-k\cdot\mathrm{trace}(M)^2 \tag{2}$$

$$M=\sum\begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix} \tag{3}$$

Where M is the self-correlation matrix of pixel intensity gradients, I_x and I_y represent horizontal and vertical gradient components, and k is an empirically determined constant.

To construct rotation-invariant descriptors, the algorithm estimates the dominant feature orientation using the intensity centroid method:

$$\theta=\tan^{-1}\left(\frac{\sum_{(x,y)\in S} y\cdot I(x,y)}{\sum_{(x,y)\in S} x\cdot I(x,y)}\right) \tag{4}$$

Where S represents pixels in the feature neighborhood, $I(x,y)$ denotes the intensity value, and (x,y) are the pixel coordinates relative to the feature point.

ORB generates binary descriptors using the BRIEF (Binary Robust Independent Elementary Features) algorithm, which randomly samples intensity values within the feature neighborhood and constructs binary feature vectors through intensity comparisons. To enhance descriptor orientation consistency, ORB applies a rotation transformation based on the estimated dominant feature direction:

$$\begin{bmatrix} x_i' \\ y_i' \end{bmatrix}=\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{5}$$

Compared to Shi-Tomasi feature detection, ORB simultaneously detects corners and generates descriptors, eliminating the need for heavy reliance on optical flow tracking. By introducing feature orientation assignment and rotation-optimized BRIEF descriptors, ORB exhibits superior robustness under high-speed UAV motion, ensuring stable and reliable feature matching.

2.2.2. Feature Booster-Based Feature Enhancement

To further enhance feature point quality, a deep learning-based Feature Booster framework is innovatively introduced. This framework improves feature descriptor performance through a dual-stage optimization mechanism, consisting of self-boosting and cross-boosting strategies.

In the self-boosting stage, a dual-path feature optimization network is constructed. The descriptor enhancement path employs a three-layer multilayer perceptron (MLP) network with nonlinear activation functions to perform feature space projection:

$$d_i^{tr}=MLP_{desc}(d_i) \tag{6}$$

Where d_i^{tr} represents the transformed feature descriptor, and MLP_{desc} is a multilayer perceptron (MLP) model designed to learn a nonlinear mapping, enhancing the robustness of descriptors during the matching process.

The geometric encoding path embeds feature point attributes into a high-dimensional space and integrates them with the self-boosted descriptors, further refining feature representations:

$$\begin{aligned} d_i^{self}&=d_i^{tr}+MLP_{geo}(p_i) \\ p_i&=(x_i,y_i,c_i,\theta_i,s_i) \end{aligned} \tag{7}$$

Where x_i and y_i represent the 2D coordinates of the feature point, θ_i represents the feature orientation angle, and c_i is the detection confidence.

In the cross-boosting stage, a lightweight Transformer architecture is introduced to perform global feature optimization. This mechanism enables each feature descriptor

to perceive the global feature distribution and interact with other feature points for improved contextual awareness. This process can be formulated as:

$$D_{cross} = Trans\left(D_{self}\right) \tag{8}$$

Where D_{cross} represents the descriptor after cross-boosting enhancement, $Trans$ denotes the Transformer model. By leveraging the self-attention mechanism, the Transformer enables feature descriptors to undergo global optimization in a high-dimensional feature space, thereby improving matching accuracy.

To enhance the matching discriminability of feature descriptors, a dual-constraint optimization objective function is designed. First, a FastAP loss term is formulated based on feature matching evaluation metrics:

$$L_{AP} = 1 - \frac{1}{N}\sum_{i=1}^{N} AP(d_i^{tr}) \tag{9}$$

Where L_{AP} represents the average precision computed using the FastAP metric, and N denotes the total number of feature points. This loss function aims to maximize feature point matching accuracy.

To further ensure the monotonic improvement of the enhancement process, an enhancement constraint term is introduced:

$$L_{BOOST} = \frac{1}{N}\sum_{i=1}^{N} max\left(0, \frac{AP(d_i)}{AP(d_i^{tr})} - 1\right) \tag{10}$$

This constraint term enforces the similarity increment between the enhanced feature and positive samples to exceed that of negative samples, ensuring the directionality of the optimization process. The final joint optimization objective is defined as:

$$L = L_{AP} + \lambda L_{BOOST} \tag{11}$$

This optimization objective ensures that after training, the Feature Booster can effectively maintain the stability of feature descriptors in complex environments.

2.2.3. RANSAC-Based Random Sample Consensus

Although the Feature Booster-enhanced feature descriptors significantly improve feature point matchability, mismatches still occur under complex imaging conditions. To mitigate the impact of outliers, which may introduce cumulative errors in pose estimation, the RANSAC (Random Sample Consensus) algorithm is introduced to eliminate incorrect matches.

The objective of RANSAC is to fit a geometric model through random sampling and to identify a set of feature points that satisfy the error constraints. In stereo camera geometry, the matching points p_{right} and p_{left} must satisfy the epipolar constraint:

$$p_{right} = H_{rans} \cdot p_{left} \tag{12}$$

Where p_{left} and p_{right} represent the matched feature points in the left and right camera views, respectively, and H_{rans} is the fundamental matrix, which describes the geometric relationship between the matched points in the stereo image pair. The reprojection error for the matched points is defined as:

$$e_{proj} = \| p_{right} - H_{rans} \cdot_{left} \|_2 \tag{13}$$

When the reprojection error is below a predefined threshold, the point is classified as an inlier; otherwise, it is considered an outlier. The core principle of RANSAC is to select the geometric model that maximizes the number of inliers. The objective function is defined as:

$$H_{opt} = \arg\max_H \sum_i \mathbb{I}(e_{proj,i} < \epsilon) \tag{14}$$

Where $\mathbb{I}(e_{proj,i} < \epsilon)$ is the indicator function, which determines whether the error is below the threshold, and $e_{proj,i}$ represents the reprojection error of the i-th matched feature point. Through random sampling and iterative optimization, RANSAC selects the model with the maximum number of inliers after multiple iterations as the final result. This approach effectively eliminates mismatched points caused by dynamic scenes or occlusions, significantly improving the accuracy and robustness of feature point matching.

2.2.4. Visual-Inertial Alignment

In visual-inertial fusion localization systems, the 3D feature coordinates and camera pose information outputted by the feature point matching and depth estimation module maintain global consistency. However, due to its low sampling rate, this data is prone to motion blur during high-speed UAV maneuvers. Conversely, the IMU sensor provides high-frequency motion parameters (above 200 Hz), but its measurement noise accumulates over time, limiting long-term localization accuracy.

To address this issue, a tightly coupled visual-inertial alignment method is adopted, establishing a unified optimization framework that integrates global constraints from visual observations with short-term motion information from the IMU.

During visual-inertial alignment, the state optimization objective within the sliding window combines the sparse global constraints from visual observations with the high-frequency short-term motion information from the IMU. Both constraints are jointly solved using a nonlinear optimization framework.

Through nonlinear optimization, the system achieves deep fusion of visual observations and inertial information in dynamic and complex environments, effectively suppressing the effects of motion blur and noise accumulation.

3. Results

To validate the applicability of the improved FB-VINS visual-inertial fusion algorithm in complex indoor environments, both simulation tests and flight experiments were conducted.

In the simulation phase, a multi-dimensional evaluation was performed using the EuRoc dataset. For the flight experiments, tests were conducted on a custom-built UAV

platform, where a motion capture system was used to collect ground truth flight data, enabling comparative analysis of localization accuracy.

3.1. UAV Platform Design

The flight verification experiments were conducted using a custom-developed small-scale quadrotor UAV platform. The UAV was equipped with multiple sensors, including an Intel RealSense D435i depth camera and a high-frequency IMU. A heterogeneous computing architecture based on the RK3588 embedded chip was implemented to enable real-time processing of both the feature enhancement algorithm and sliding window optimization.

The system adopted a modular communication architecture, facilitating high-speed data transmission between the onboard computer, sensor modules, and flight control module through a combination of serial buses and parallel interfaces.

The UAV utilized the PX4 flight control system, with QGroundControl ground station software employed for flight initialization and parameter configuration. Before the experiments, high-precision pose data from the motion capture system was used to fine-tune the UAV's PID parameters, ensuring stable and controlled flight in complex environments. The developed UAV platform is illustrated in Figure 2.

Figure 2. UAV Structural Schematic Diagram

3.2. EuRoc Dataset Simulation Experiments

To evaluate the localization capability of the improved visual-inertial fusion algorithm (FB-VINS) in complex indoor environments, a simulation experiment was conducted using the EuRoc MAV open-source dataset. The EuRoc dataset provides stereo image sequences, IMU sensor data, and ground truth trajectories, making it a widely used benchmark for visual-inertial odometry (VIO) and SLAM algorithms.

The experiment selected six representative sequences from the EuRoc dataset to comprehensively assess FB-VINS's localization performance across different scenarios. MH_01_easy and MH_03_medium were captured in industrial environments containing robotic arms, pipelines, and metal structures. These sequences feature rich textures but dim lighting, making them suitable for evaluating localization accuracy in low-light and structured environments. V1_01_easy and V1_02_medium include significant illumination variations and sparse texture regions, testing the algorithm's robustness to extreme lighting changes. V2_01_easy and V2_02_medium feature furniture such as tables, chairs, and bookshelves, with sparse textures, dynamic occlusions, and local feature loss, providing a benchmark for assessing adaptability in dynamic environments with occlusions. Through these diverse experimental scenarios, the localization capability of FB-VINS in complex indoor environments was systematically validated.

To quantitatively evaluate performance, a comparative analysis was conducted between FB-VINS and the original VINS-Fusion algorithm, measuring localization errors across different sequences. The absolute root mean square error (RMSE) was used as the primary evaluation metric, while additional statistical indicators including maximum error (Max), mean error (Mean), median error (Median), and standard deviation error (Std) were analyzed to provide a comprehensive assessment of localization accuracy.

As shown in Table 1, FB-VINS exhibited improvements over VINS-Fusion across all error metrics. The maximum error (Max) was reduced by 3.28%, the mean error (Mean) decreased by 12.89%, and the median error (Median) dropped by 13.67%. Additionally, the root mean square error (RMSE) improved by 12.24%, and the standard deviation error (Std) decreased by 9.70%. These results indicate that the FB-VINS algorithm consistently reduces localization errors, with particularly notable improvements in RMSE, highlighting its enhanced robustness and accuracy in complex indoor environments.

Table 1. Lateral Quantitative Comparison on the EuRoc Dataset

sequence	algorithm	Max	Mean	Medium	RMSE	Std
MH_01_easy	FB-VINS	0.6649	0.3805	0.3879	0.3982	0.1174
MH_01_easy	VINS-Fusion	0.7920	0.4221	0.4604	0.4486	0.1519
MH_03_medium	FB-VINS	0.9541	0.3724	0.3098	0.4223	0.1992
MH_03_medium	VINS-Fusion	1.0342	0.4455	0.3724	0.5068	0.2416
V1_01_easy	FB-VINS	0.2064	0.1047	0.1013	0.1123	0.0401
V1_01_easy	VINS-Fusion	0.3080	0.1215	0.1151	0.1319	0.0513
V1_02_medium	FB-VINS	0.2259	0.1006	0.0956	0.1082	0.0401
V1_02_medium	VINS-Fusion	0.2128	0.1161	0.1131	0.1222	0.0381
V2_01_easy	FB-VINS	0.2034	0.0786	0.0709	0.0853	0.0331
V2_01_easy	VINS-Fusion	0.1688	0.0857	0.0806	0.0911	0.0309
V2_02_medium	FB-VINS	0.5169	0.0839	0.0651	0.1087	0.0690
V2_02_medium	VINS-Fusion	0.4681	0.0994	0.0723	0.1247	0.0754

3.3. UAV Flight Verification

To further validate the adaptability and localization accuracy of the improved visual-inertial fusion algorithm in real-world environments, a UAV flight experiment was conducted in an indoor motion capture (MoCap) test facility. The facility was equipped with an optical motion capture system, which provided high-precision pose reference trajectories, enabling accurate assessment of the algorithm's

real-world localization errors.

The experiment required the UAV to perform a stationary waypoint flight task with a designated target point at (0.6, 0.3, 0.6) meters. The UAV took off from (0, 0, 0.6) and sequentially passed through waypoints (0.6, 0, 0.6) before reaching the final target position. Throughout the flight, programmatic flight control was used instead of manual operation to eliminate human interference and ensure repeatable testing conditions.

The motion capture system continuously recorded the UAV's real-time position data, and a comparative analysis was performed to evaluate the axis-wise position errors of the algorithm before and after improvement. The trajectory recorded by the motion capture system during the UAV flight experiment is illustrated in Figure 3.

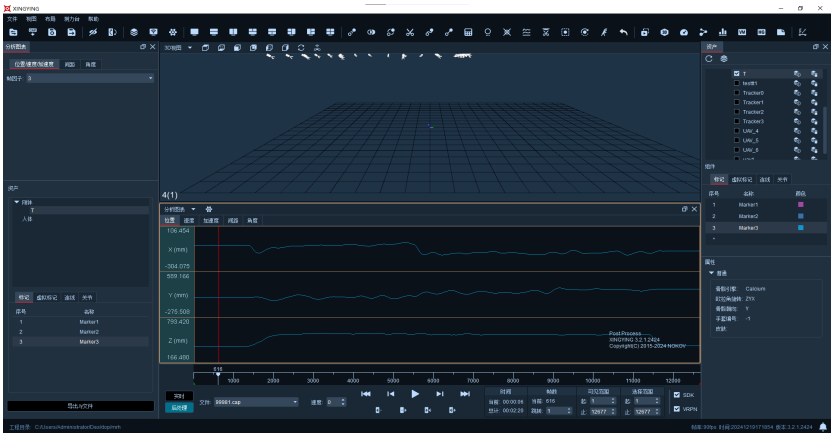

Figure 3. UAV Trajectory Visualization

Before the flight experiment, motion capture markers were attached to the UAV, and a rigid body model was constructed. Due to variations in UAV height and marker placement, the measured initial position did not coincide with the origin. During the flight, a high-resolution camera system continuously recorded and transmitted real-time UAV pose information for analysis.

The experiment recorded and compared the autonomous localization performance of the UAV using both the traditional VINS-Fusion algorithm and the improved FB-VINS algorithm. The UAV's actual stabilized position after reaching the target point was compared with the expected position, providing a quantitative evaluation of the accuracy and feasibility of the improved algorithm in real-world applications. The final recorded results are presented in Tables 2 and 3. Based on the experimental analysis, the UAV's axis-wise positioning errors were significantly reduced when using the FB-VINS algorithm compared to the original VINS-Fusion algorithm. This improvement demonstrates enhanced localization accuracy, verifying the effectiveness and practical applicability of the proposed algorithm in autonomous UAV positioning.

Table 2. Motion Capture-Based Fixed-Point Flight Results Using the VINS-Fusion Algorithm

	starting-point measured value	Ending-point measured value	single-axis absolute positional difference	expected value	axial error
X-axis	-9.08cm	55.36cm	64.44cm	60cm	4.40cm
Y-axis	-1.18cm	23.30cm	24.48cm	30cm	5.52cm
Z-axis	17.65cm	73.10cm	55.45cm	60cm	4.55cm

Table 3. Motion Capture-Based Fixed-Point Flight Results Using the FB-VINS Algorithm

	starting-point measured value	Ending-point measured value	single-axis absolute positional difference	expected value	axial error
X-axis	-9.37cm	53.12cm	62.49cm	60cm	2.49cm
Y-axis	-1.56cm	23.80cm	25.36cm	30cm	4.64cm
Z-axis	17.65cm	74.51cm	56.86cm	60cm	3.14cm

4. Conclusions

This paper addresses the limitations of VINS-Fusion in feature extraction and matching robustness under weak-texture, dynamic lighting, and partial occlusion environments. To overcome these challenges, an improved visual-inertial fusion algorithm (FB-VINS) is proposed, incorporating ORB-based feature extraction and Feature Booster-based feature enhancement.

In the visual frontend, ORB replaces the original Shi-Tomasi feature detector, and a deep learning-driven feature enhancement technique is integrated to optimize feature descriptor quality. This improves feature stability in complex environments, enhancing the system's overall localization reliability.

To validate the algorithm's effectiveness, simulation experiments were conducted using the EuRoc dataset, followed by real-world flight tests on a quadrotor UAV

platform. Experimental results demonstrate the feasibility and effectiveness of the proposed visual odometry improvement method. Compared to VINS-Fusion, FB-VINS achieves higher-quality feature point detection and more accurate trajectory estimation in challenging environments.

By improving localization accuracy, FB-VINS also enhances system robustness and adaptability, making it more suitable for indoor UAV navigation in complex and GNSS-denied environments.

Acknowledgment

I would first like to express my sincere gratitude to my supervisor, Professor Xi Zhang, Jun Qi, and Weidong Peng, for their invaluable guidance in shaping the research questions and methodology. Their insightful feedback continuously challenged me to refine my ideas and significantly enhanced the quality of this work.

Additionally, I am deeply grateful to my friends for their unwavering support, engaging discussions, and the much-needed moments of respite, which provided balance and motivation throughout this research journey. Finally, | feel indebted to my parents, for their understanding and love in life.

References

[1] X. Y. Wang, G. Z. Gao, L. L. Gu, et al., “Study on the legal system for promoting China’s low-altitude economy,” Information & Communication Technology and Policy, no. 11, pp. 48–53, 2024.

[2] T. Xia and J. L. Zhao, “Exploration on the application of UAV technology in military logistics,” China Storage and Transportation, no. 9, pp. 132–133, 2024.

[3] L. Yang, “Analysis of the effectiveness of water conservancy engineering inspection using UAV technology,” Water Safety, no. 21, pp. 13–15, 2024.

[4] X. W. Cui, “Application of UAV technology in geological disaster monitoring,” Anhui Geology, no. 2, pp. 153–156, 2024.

[5] Z. H. Wang, “Indoor positioning and flight control of quadrotor UAV based on UWB,” M.S. thesis, Heilongjiang Univ., Harbin, China, 2024.

[6] M. Zhao, “Study on UAV autonomous positioning method based on 3D LiDAR,” M.S. thesis, North China Electric Power Univ., Beijing, China, 2020.

[7] F. Y. Li, “Design and implementation of UAV indoor positioning system based on UWB technology,” M.S. thesis, Nanchang Univ., Nanchang, China, 2022.

[8] J. F. He, T. Xi, and L. Zhang, “Design of a rotor UAV positioning system based on monocular vision,” Wireless Internet Technol. no. 21, pp. 79–81, 2022.

[9] B. Chen and Q. Li, “Application of multi-sensor fusion in UAV indoor 3D positioning,” Sensor World, no. 3, pp. 27–36, 2022.

[10] L. Y. Wang, G. T. Yang, B. Y. Li, et al., “UAV indoor positioning and autonomous navigation method based on improved VINS,” Control Eng., to be published.

[11] Y. F. Cai, Z. H. Lu, Y. C. Li, et al., “A tightly coupled SLAM system based on multi-sensor fusion,” Automot. Eng., no. 3, pp. 350–361, 2022.

[12] J. C. Li, Z. C. Zhang, B. Y. Ding, et al., “Research on visual-inertial SLAM based on small sliding window optimization,” Electronic Production, no. 22, pp. 96–99, 2022.

[13] X. Wang, Z. Liu, Y. Hu, et al., “FeatureBooster: Boosting feature descriptors with a lightweight neural network,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, New York, NY, USA, 2023, pp. 7630–7639.

[14] W. W. Zhang, J. L. Zhao, J. He, et al., “A step array-based BRIEF feature descriptor,” Computer Technol. Dev., no. 5, pp. 81–87, 2023.

[15] S. Q. Tan and P. J. Shuai, “Research on ORB feature point extraction algorithm based on adaptive threshold,” Mod. Comput., no. 4, pp. 43–47, 2024.

[16] X. J. Ning, J. R. Li, F. Gao, et al., “Feature matching algorithm based on optimal geometric constraints and RANSAC,” J. Syst. Simul., no. 4, pp. 727–734, 2022.

[17] J. H. Lu, “Research on multi-sensor fusion SLAM based on tight coupling and graph optimization,” M.S. thesis, Chongqing Technol. and Business Univ., Chongqing, China, 2023.

[18] W. S. Chen, Y. F. Huang, and X. F. Lu, “Review of UAV detection technology based on multi-sensor fusion,” Mod. Radar, no. 6, pp. 15–29, 2020.

[19] J. W. Jiang, Y. H. Ji, J. J. Liu, et al., “A fast loop detection algorithm in visual SLAM based on lightweight CNN,” Comput. Simul., no. 8, pp. 182–188, 2024.

[20] Z. G. Shang, “Research on visual SLAM algorithm based on particle swarm optimization and deep learning,” M.S. thesis, Univ. of Electronic Sci. and Technol. of China, Chengdu, China, 2024.