Structure Information Aided Visual/Inertial Odometry of UAVs for Unidirectional Structural Environment

XIE Shenghong[1], *WEN Yebei*[1], *WANG Hongfeng*[2], *HUANG Zheng*[1], *JI Tianyang*[1], *YUAN Cheng*[1], *LYU Pin*[1*]

1. College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, P. R. China;
2. Beijing Institute of Automation Control Equipment, Beijing 100074, P. R. China

(Received 10 September 2025; revised 26 February 2026; accepted 16 March 2026)

Abstract: The integrated utility system is indispensable for maintaining urban stability and services. Regular unmanned aerial vehicles (UAVs) inspections of these facilities are critical for ensuring operational urban systems. Such scenarios, characterized by spatially confined layouts and unidirectional extension, are defined as unidirectional structural (UDS) environment. Given the sparse textural features inherent and darker light to UDS environment, conventional methods fail to ensure safe UAV navigation. Therefore, this paper proposes a structure-assisted visual-inertial odometry algorithm for UAVs in dark UDS environment. The approach enhances the accuracy of camera feature extraction through an adaptive median filtering and BM3D (AMF-BM3D) denoising method, and incorporates an inertial fault-tolerant vanishing point technique to achieve high-precision, cumulative-error-free UAV yaw estimation, while utilizing gravity constraints to eliminate cumulative errors in roll and pitch angles. Finally, the extended Kalman filter (EKF) fuses pose data from conventional visual-inertial odometry (VIO) to obtain refined pose estimation. Experimental validation in UDS environment demonstrates that compared to traditional VIO, the algorithm proposed significantly enhances the accuracy of pose estimation and the safety of UAV flight in the UDS environment.

Key words: structure information; visual-inertial odometry (VIO); extended kalman filter (EKF)

CLC number: V249.32 **Document code:** A **Article ID:** 1005-1120(2026)04-0546-14

0 Introduction

Nowadays, unmanned aerial vehicles[1] (UAVs) are extensively utilized in inspection, rescue operations, aerial photography and various other fields due to their high flexibility, mobility, adaptability and other attributes[2]. UAVs equipped with cameras and other sensors can navigate autonomously[3-5], and thus operate effectively in challenging environments. Harsh working environments, such as tunnels, chimneys and boilers, are characterized by complexity, airtightness, satellite signal obstruction and other factors.

The unidirectional structural (UDS) environment is frequently encountered, such as corridors, tunnels and power pipe corridors. It is characterized by its airtightness, extensive longitudinal distances, narrow spaces, unidirectional extension and poor illumination. Underground utility corridors, which integrate power, telecommunications, gas, heating and water supply pipelines, are critical urban infrastructures. They currently rely on manual inspections, which are costly[6] and unsafe.

In recent years, numerous scholars and institutions have conducted extensive research on UAV navigation in UDS environments. To address the challenges of navigation in these environments, vision-based methods have been extensively researched and deployed. Compared to active sensors

*Corresponding author, E-mail address: lvpin@nuaa.edu.cn.

How to cite this article: XIE Shenghong, WEN Yebei, WANG Hongfeng, et al. Structure information aided visual/inertial odometry of UAVs for unidirectional structural environment[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2026, 43(4): 546-559.

http://dx.doi.org/10.16356/j.1005-1120.2026.04.005

like LiDAR, visual sensors offer significant advantages in terms of size, weight and power, as well as cost-effectiveness[7]. Furthermore, cameras provide rich texture and semantic information, which are benificial for scene understanding and long-term operation[8].

Benefiting from the capability to operate without reliance on prior maps or external infrastructure, visual simultaneous localization and mapping (V-SLAM) and visual-inertial odometry (VIO) have become the mainstream solutions for UAV autonomous navigation. In particular, VIO, which fuses visual data with high-frequency inertial measurement unit (IMU) measurements, effectively compensates for the lack of scale in monocular vision and improves robustness against rapid movements and temporary visual tracking failures[9-10]. Thus, VIO-based methods have been widely used to achieve precise state estimation for UAVs in complex scenarios.

However, the performance of VIO systems is fundamentally determined by the quality of the input data. Image acquisition serves as the front-end of the VIO pipeline. Accurate, reliable image acquisition is the prerequisite for ensuring localization precision. In UDS environments characterized by low illumination and dynamic lighting changes, poor image quality can directly lead to feature tracking failures and significant drift in navigation[11].

In previous studies on poor illumination images, Gaussian noise has included the main noise source. Image denoising methods are usually divided into spatial domain methods and frequency domain methods[12]. In recent years, the BM3D[13] algorithm has been widely used due to its excellent performance in removing Gaussian noise. This method groups similar image blocks and performs collaborative filtering in three dimensions, achieving good denoising effects while preserving details. However, BM3D is sensitive to non-Gaussian noise such as salt-and-pepper noise and performs poorly in such cases, especially when isolated noise points appear in low-light nighttime images, making it difficult to remove noise effectively and affecting image quality and subsequent feature extraction[14]. This necessitates a robust preprocessing front-end capable of handling mixed noise to ensure the stability of the VIO system.

Beyond image quality, the estimation backend faces inherent challenges in UDS environments. Mainstream VIO algorithms fall into two categories, filter-based methods[15] and optimization-based methods[16]. Although filter-based methods are computationally efficient, they are prone to linearization errors. Conversely, optimization-based methods achieve higher accuracy but consume more resources. Regardless of the architecture, all dead-reckoning systems inevitably suffer from cumulative error over time due to sensor noise integration and bias estimation inaccuracies.

To mitigate error accumulation, Li et al.[17], Elmokadem et al.[18] and Wang et al.[19] employed visual perception approaches for UAV pose estimation. However, these methods struggle with the feature-poor and dimly lit UDS corridors. The standard solution to correct drift is loop closure detection. Gao et al.[20] introduced feature-matching loop closure, while Wen et al.[21] utilized semantic topology maps. Wang et al.[22] and Sun et al.[23] proposed distance-based detection to reset pose errors. However, the UDS environment presents a unique unidirectional extension characteristic. The UAV flies long distances without returning to previous locations, meaning the geometric condition for loop closure is inherently absent. Consequently, the system cannot eliminate cumulative errors through global optimization, presenting a severe challenge for long-endurance inspection.

With the advancement of artificial intelligence, learning-based approaches offer novel solutions to address the challenges of robust feature representation and complex motion dynamics in VIO. Clark[24] proposed VINet, modeling VIO as a sequence-to-sequence regression problem using long short-term memory (LSTM) to address long-term dependencies. Similarly, DeepVO[25] and UnDeepVO[26] leverage convolutional neural networks (CNNs) to estimate pose directly from raw images, by passing traditional feature engineering.

However, recent surveys indicate that deep learning-based VIO still faces challenges in generalization across different domains and deployment on resource-constrained edge devices[27]. First, deep learning models heavily rely on the distribution of training data. Models trained on standard datasets perform poorly in UDS environments due to the significant domain gap in lighting and texture. Second, most open-source models are based on the pinhole camera model assumption. For inspection UAVs equipped with fisheye cameras to capture wider fields of view, the severe distortion requires specialized network redesign and retraining, which is data-intensive. Finally, deep neural networks require substantial computing power. Given the strict size, weight and power constraints of edge computing modules on small UAVs, running heavy inference models will compromise flight endurance and real-time performance.

Addressing the limitations of existing algorithms, this paper proposes a lightweight, robust navigation system tailored for UDS environments. We propose a VIO system based on structural regularity assistance. By leveraging the parallel structural features inherent in UDS environments as global constraints, we correct the heading and position drift without relying on loop closure. Furthermore, to ensure high-quality input for the VIO backend, we introduce an adaptive median filtering and BM3D (AMF-BM3D) method. This hybrid denoising strategy effectively eliminates mixed noise in low-light fisheye images, significantly improving feature extraction accuracy.

The main contributions are as follows.

(1) We introduce a binary mask to detect the presence of salt-and-pepper noise in the image. For pixels identified as contaminated by salt-and-pepper noise, median filtering is applied as a preprocessing step, followed by denoising using the BM3D algorithm.

(2) We present a relative yaw estimation algorithm based on the inertial fault-tolerant vanishing point method. The relative yaw is estimated following the Manhattan assumption within UDS environment, with short-term inertial recursive information utilized to sustain yaw stability amid sudden environmental changes. This algorithm enhances the robustness of relative yaw estimation, mitigating error accumulation over time.

(3) We introduce a roll and pitch algorithm that relies on gravity constraints and complementary filtering. The roll and pitch angles are constrained by gravity, effectively minimizing drift.

Experimental validation demonstrates that the proposed structure-assisted visual-inertial pose estimation algorithm significantly enhances attitude angle estimation accuracy compared to conventional VIO approaches.

The main structure of this paper is as follows. Section 1 introduces the visual-inertial odometry system based on structure information assistance. Section 2 introduces the specific methods proposed in this paper, including relative yaw estimation utilizing the vanishing point method, and roll and pitch estimation employing complementary filtering. Section 3 presents the experimental details and results in the UDS environment. Finally, Section 4 presents the conclusion.

1 Overview

1.1 Framework

The architecture of the visual-inertial odometry system proposed in this paper is illustrated in Fig.1. It mainly consists of five modules: A low-light image denoising module based on AMF-BM3D, an optical flow visual odometry module, a relative yaw estimation module based on an inertial fault-tolerant vanishing point method, a roll and pitch estimation module based on gravity-constrained complementary filtering and a pose optimization module based on extended Kalman filter (EKF).

The AMF-BM3D module preprocesses images captured by the fisheye camera to remove noise and increase the proportion of valid feature points, thus laying a solid foundation for feature extraction. Visual odometry based on the optical flow method provides a preliminary pose estimation of UAVs in UDS environment (Section 2.1). We utilize the

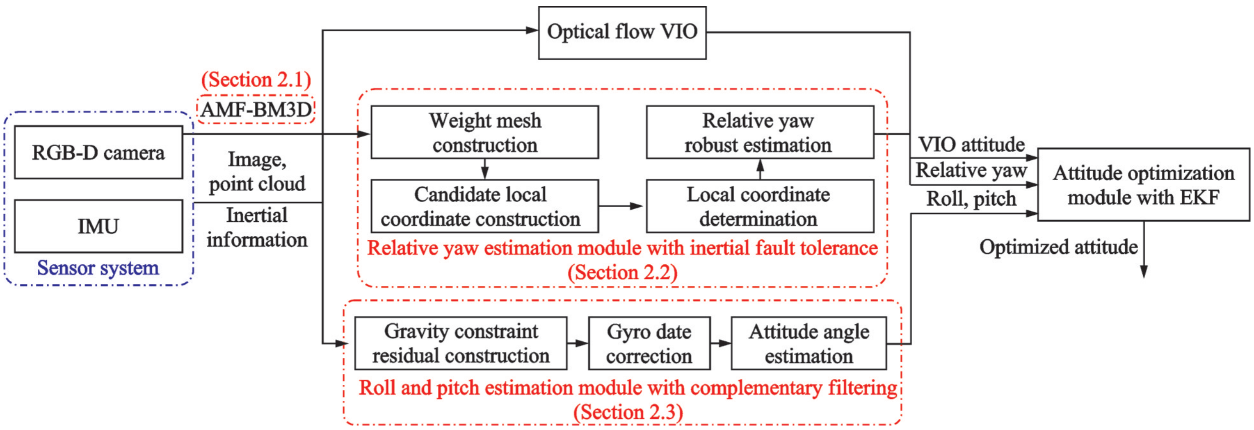


Fig.1 Visual-inertial odometry system

vanishing point method, based on the Manhattan world (MW) assumption, to construct the Manhattan coordinate system and determine the relative yaw of the UAV (Section 2.2). Complementary filtering is employed to estimate the roll and pitch of the UAV using IMU data (Section 2.3). Finally, EKF is utilized to fuse the relative yaw, roll, and pitch measurements, resulting in a more accurate UAV pose estimation.

1.2 Notation

In this paper, the world coordinate system is defined as $W=\{\hat{x}_W, \hat{y}_W, \hat{z}_W\}$, where $\hat{x}_W$ denotes the direction along the central axis of the UDS environment and $\hat{z}_W$ the direction opposite to the gravity. The body coordinate system is defined as $B=\{\hat{x}_B, \hat{y}_B, \hat{z}_B\}$. At the initial time, when the yaw angle of the UAV is 0°, the body coordinate system aligns with the world coordinate system.

As shown in Fig.2, utilizing the uniaxial and symmetrical characteristics of the UDS environment, we establish the local coordinate system denoted as $L=\{\hat{x}_L, \hat{y}_L, \hat{z}_L\}$. $\hat{x}_L$ represents the tangent direction pointing to the central axis of the UDS environment, $\hat{y}_L$ the normal direction of the plane parallel to both sides of the UDS environment and $\hat{z}_L$ the direction opposite to gravity. The black dashed line represents the central axis of the UDS environment, and the blue dashed line represents the flight trajectory of the UAV. The origin of this local coordinate system consistently resides on the central axis of the UDS environment.

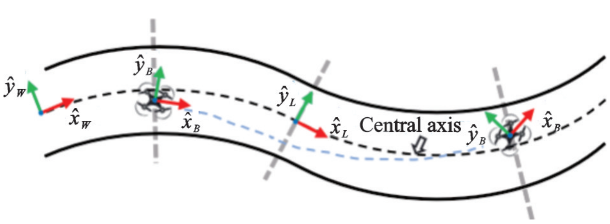


Fig.2 Top view of quadrotor UAV flying in UDS environment

In this paper, we define the rotation matrix as $\boldsymbol{R}_1^2$ indicating the transformation of a given point from coordinate system 1 to coordinate system 2.

2 Methodology

2.1 Adaptive median filtering and BM3D denoising

In low-light corridor environments, the image quality captured by fisheye cameras is often affected by many aspects. Due to the weak signal of the sensor in low light conditions in the imaging principle, the signal-to-noise ratio is significantly reduced, which in turn increases the number of Gaussian noise and hot pixels. Further, in the dedistortion process of fisheye cameras, the interpolation algorithm needs to be calculated based on the pixels of the original image, and the hot pixels in the original image are amplified or diffused in the interpolation calculation, resulting in the generation of random and isolated salt-and-pepper noise. As shown in Fig.3, the noise effect of dark fisheye images reduces the number of effective feature points, the quality of descriptors and the accuracy of feature extraction.

Various denoising methods have been developed to date. The widely used BM3D algorithm[22] excellently performs in suppressing Gaussian noise but is sensitive to nonlinear noise such as salt-and-pepper noise, resulting in suboptimal noise removal. Building on this, we propose AMF-BM3D based on salt-and-pepper noise detection. The process and effect are shown in Fig.3.

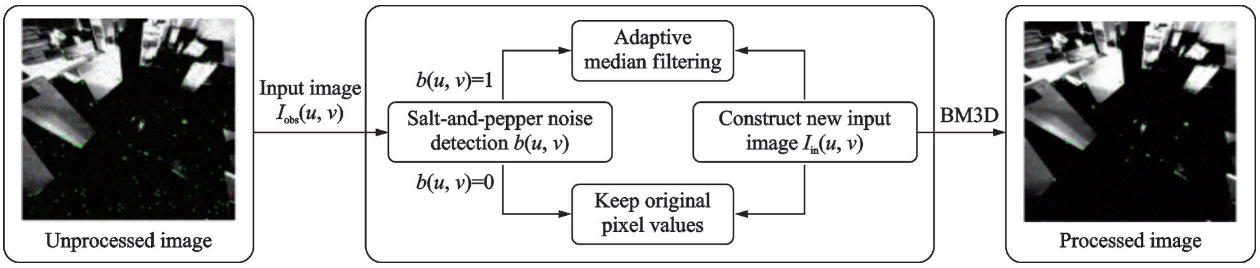


Fig.3 Flow chart and renderings of AMF-BM3D

The noisy image is modeled as

$$I_{obs}(u,v)=I_{clean}(u,v)+\eta(u,v)+b(u,v)s(u,v) \tag{1}$$

where $I(u, v)$ represents the pixel coordinates in the image plane, with u and v representing the horizontal and vertical coordinates, respectively. The observed noisy image and the clean image is represented as $I_{obs}(u, v)$ and $I_{clean}(u, v)$. $\eta(u, v)\sim N(0, \sigma^2)$ represents Gaussian noise, $b(u, v)$ is a binary mask indicating whether pixel at (u, v) is contaminated by salt-and-pepper noise, and $s(u, v)$ is the salt-and-pepper noise function.

To accurately identify salt-and-pepper noise pixels, the detection function is defined as

$$b(u,v)=\begin{cases}1 & I_{obs}(u,v)=I_{min},I_{max}\\ 0 & \text{Otherwise}\end{cases} \tag{2}$$

where I_{min} and I_{max} represent the minimum and maximum grayscale values in the image.

For pixels detected as salt-and-pepper noise, median filtering is applied as a preprocessing step to effectively suppress extreme noise values that may affect subsequent filtering. The median filtering operation is defined as

$$I_{med}(u,v)=\text{med}\{I_{obs}(i,j)\,|\,(i,j)\in \boldsymbol{W}_{u,v}\} \tag{3}$$

where $\boldsymbol{W}_{u,v}$ is an $n\times n$ local neighborhood window centered at pixel (u, v), and (i, j) indexes the pixels within this neighborhood.

The input image for BM3D is construced based on $b(u, v)$

$$I_{in}(u,v)=(1-b(u,v))\cdot I_{obs}(u,v)+b(u,v)\cdot I_{med}(u,v) \tag{4}$$

The pixels affected by salt-and-pepper noise is replaced with the results of median filtering, while keeping the original values are maintained for the non-salt-and-pepper noise pixels. Then, the classic BM3D algorithm is used to the image $I_{in}(u, v)$ for denoising to obtain the final estimated image. As shown in Fig.3, the proposed method effectively suppresses salt-and-pepper noise in dark fisheye images and improves the robustness and accuracy of feature point extraction.

2.2 Relative yaw estimation based on inertial fault tolerance and vanishing point

Given the abundant structural features present in the UDS environment, the MW hypothesis is employed to establish the local coordinate system and accurately estimate the UAV's relative yaw, ensuring its stability over time. Nevertheless, conventional estimation methods remain susceptible to environmental variations, where inaccurate relative yaw estimation can significantly impact the pose estimation of UAVs. Hence, this section introduces an inertial fault tolerance-based method to maintain consistent and stable relative yaw under dynamic environmental conditions. The specific process of UAV relative yaw estimation is as follows.

2.2.1 Local coordinate estimation

Based on the Manhattan world assumption[28], we propose a method for estimating a local coordinate system in the UDS environment. First, a unit sphere is constructed, centered on the camera focus. In the unit sphere, the camera coordinate system $C=\{\hat{x}_C, \hat{y}_C, \hat{z}_C\}$ is established, with the sphere's center serving as the origin. In this paper, the sharp bracket is used to define the coordinate system, which is the ordered set of three orthogonal unit vectors. The X axis and Y axis align with the u axis and v axis of the image, respectively. The Z axis of the coordinate system extends from the camera's focus to the center of the image $(u_c, v_c)^T$. The coordinates under the camera coordinate system are expressed as

$$\begin{cases} x_C = u - u_c \\ y_C = v - v_c \\ z_C = f \end{cases} \tag{5}$$

As shown in Fig.4, a polar coordinate system $P = \{\phi, \lambda\}$ is established on the image plane, where ϕ and λ are expressed in radians, with $\phi \in [0, \pi/2]$ and $\lambda \in [0, 2\pi)$. The coordinates of any point $(x_C, y_C, z_C)^T$ within the camera coordinate system under the polar coordinate system are expressed as

$$\begin{cases} \phi = \arccos\left(\dfrac{z_C}{\sqrt{(x_C)^2 + (y_C)^2 + (z_C)^2}}\right) \\ \lambda = \arctan\left(\dfrac{x_C}{y_C}\right) + \pi \end{cases} \tag{6}$$

Subsequently, the angular domain is uniformly discretized with a resolution of 1°, resulting in a 90 × 360 weight grid. Let $i_\phi = 1, 2, \cdots 90$, and $i_\lambda = 1, 2, \cdots 360$ denote the grid indices in the ϕ-directions and λ-directions, respectively. The weight of each grid is denoted by $\psi(i_\phi, i_\lambda)$. To extract line features from a given image, we employ the line segment detector (LSD) algorithm. The intersection points of two image line features fall on the grid of the weight grid graph. If the number of intersection points falling on the grid is $N_{(\phi_{\deg}, \lambda_{\deg})}$, the weight assigned to the grid is expressed as

$$\psi(\phi_{\deg}, \lambda_{\deg}) = \sum_{i=1}^{N_{(\phi_{\deg}, \lambda_{\deg})}} (\| s_1 \| \cdot \| s_2 \| \cdot \sin(2\theta)) \tag{7}$$

where $\| s_1 \|$ and $\| s_2 \|$ denote the lengths of the two line segments associated with the l-th intersection point, and θ denotes the included angle between these two segments.

As shown in Fig.4, we construct a candidate local coordinate system. In this paper, a line feature intersection point is arbitrarily selected in the polar coordinate system, and its corresponding point A in the camera coordinate system is found. The vector $\overrightarrow{OA}$ is the first coordinate axis $\boldsymbol{V}_1$ of the candidate local coordinate system. Along the intersection line between the plane perpendicular to $\boldsymbol{V}_1$ and the unit sphere, points are taken at intervals of 1° to form the second coordinate axis $\boldsymbol{V}_2$ of the candidate local coordinate system. The third coordinate axis $\boldsymbol{V}_3$ needs to meet $\boldsymbol{V}_3 = \boldsymbol{V}_1 \times \boldsymbol{V}_2$.

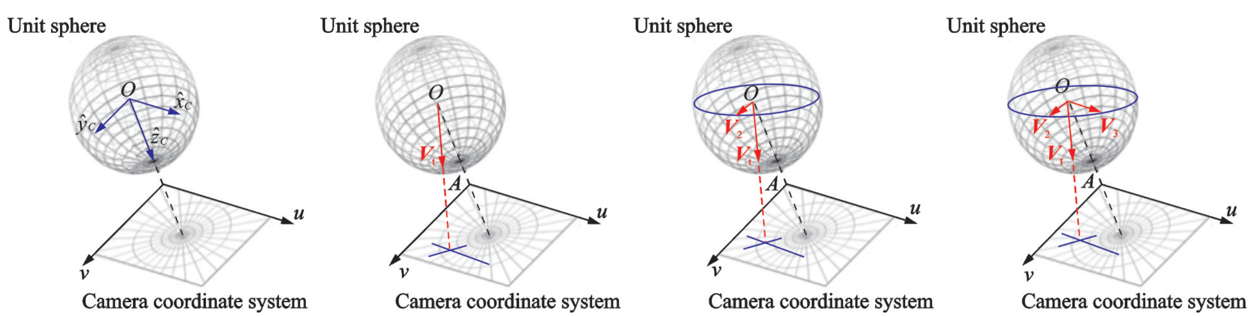


Fig.4 Construction process of candidate local coordinate systems

Finally, the most suitable local coordinate system is selected from the candidate local coordinate systems by using the weight grid graph. The three axes of the candidate local coordinate system are projected into the polar coordinate system. The grid weight value is retrieved and the weight ξ of the candidate coordinate system is constructed as

$$\xi = \psi(\phi_1, \lambda_1) + \psi(\phi_2, \lambda_2) + \psi(\phi_3, \lambda_3) \tag{8}$$

We select the local coordinate system with the highest weight among all candidate local coordinate systems.

To enhance the accuracy of the local coordinate system estimation, a ground constraint is introduced to validate the selected local coordinate system. Given a ground normal vector $\boldsymbol{n}_d = \{x_d, y_d, z_d\}^T$, axis of the local coordinate system is obtained

$$\partial = \arccos\left(\frac{\boldsymbol{n}_{\mathrm{d}} \cdot \hat{z}_L}{\|\boldsymbol{n}_{\mathrm{d}}\| \cdot \|\hat{z}_L\|}\right) \tag{9}$$

When the included angle is less than the ground constraint threshold, denoted as σ_{d}, the local coordinate system is considered to be accurate. Otherwise, if the estimation of the local coordinate system is considered to be inaccurate, it must be reestimated.

2.2.2 Robust estimation of relative attitude angle

According to the local coordinate system, we can estimate the attitude of UAV relative to the environment, denoted as $\boldsymbol{\alpha}_{\mathrm{Angle}} = (\alpha_{\mathrm{roll}}, \alpha_{\mathrm{pitch}}, \alpha_{\mathrm{yaw}})^{\mathrm{T}}$. We introduce an outlier detection algorithm based on residual analysis, leveraging IMU data, to enhance the robustness of relative attitude angle estimation.

Firstly, we utilize IMU data to calculate the change in UAV relative attitude within the interval between two estimates of the local coordinate system, which is recorded as $\boldsymbol{\alpha}_{\mathrm{Angle}} = (\alpha_{\mathrm{roll}}, \alpha_{\mathrm{pitch}}, \alpha_{\mathrm{yaw}})^{\mathrm{T}}$. The relative attitude at $k-1$ moment is $\boldsymbol{\alpha}_{\mathrm{Angle}_{k-1}^r}$. Then the predicted relative attitude at k moment can be expressed as

$$\boldsymbol{\alpha}_{\widetilde{\mathrm{Angle}_k^r}} = \boldsymbol{\alpha}_{\mathrm{Angle}_{k-1}^r} + \Delta\boldsymbol{\alpha}_{\mathrm{Angle}} \tag{10}$$

Subsequently, we construct absolute residua Ω_{abs} and relative residua Ω_{rel} according to the relative attitude $\alpha_{\mathrm{Angle}_k^r}$ estimated at time k

$$\begin{cases} \Omega_{\mathrm{abs}} = \left| \boldsymbol{\alpha}_{\mathrm{Angle}_k^r} - \boldsymbol{\alpha}_{\widetilde{\mathrm{Angle}_k^r}} \right| \\ \Omega_{\mathrm{rel}} = \dfrac{\left(\alpha_{\mathrm{roll}_k^r} - \alpha_{\widetilde{\mathrm{roll}_k^r}}\right)^2}{\alpha_{\widetilde{\mathrm{roll}_k^r}}} + \dfrac{\left(\alpha_{\mathrm{pitch}_k^r} - \alpha_{\widetilde{\mathrm{pitch}_k^r}}\right)^2}{\alpha_{\widetilde{\mathrm{pitch}_k^r}}} + \dfrac{\left(\alpha_{\mathrm{yaw}_k^r} - \alpha_{\widetilde{\mathrm{yaw}_k^r}}\right)^2}{\alpha_{\widetilde{\mathrm{yaw}_k^r}}} \end{cases} \tag{11}$$

When Ω_{abs} and Ω_{rel} meet any of the formulas in Eq.(12), it is concluded that there exists a significant error in the relative attitude angle.

$$\begin{cases} \Omega_{\mathrm{abs}} > \sigma_{\mathrm{static}} + n \cdot \sigma_{\mathrm{dynamic}} \\ \Omega_{\mathrm{rel}} > \sigma_{\mathrm{rel}} \end{cases} \tag{12}$$

where σ_{static} serves as the static baseline, representing the minimum permissible error range between the IMU prediction and visual observation; n denotes the count of consecutive errors, when the visual system fails upon entering a dark region, and n increments by 1 each time when the system rejects visual data; $\sigma_{\mathrm{dynamic}}$ is a compensation coefficient correlated with the gyroscope's random walk; and σ_{rel} is a fixed threshold, if the change in visual angle between adjacent frames exceeds this value, and the system identifies it as an abnormal outlier. Its specific value depends on the maximum angular velocity of the UAV and the frame rate of the camera.

Finally, for the relative attitude angle with significant estimation error, the predicted value $\boldsymbol{\alpha}_{\widetilde{\mathrm{Angle}_k^r}}$ replaces the estimated result $\boldsymbol{\alpha}_{\mathrm{Angle}_k^r}$ to determine the relative α_{yaw^r} of the UAV. Considering that the predicted value based on IMU recursion will diverge over time, we introduce a dynamic threshold based on IMU gyro random walk in determining the absolute residual. As the number of error estimation instances for relative attitude angles increases, the range of the absolute residual judgment threshold expands, thereby enhancing the robustness of residual judgment.

2.3 Roll and pitch estimation based on complementary filtering

The estimation of the roll and pitch of the UAV from VIO will diverge over time. According to the characteristics of small maneuvers and the slow speed of UAVs in an unknown UDS environment, we introduce a gravity constraint and proposes a method for estimating roll and pitch based on complementary filtering, which improves the accuracy of estimation for roll and pitch.

The flight of UAVs in the unknown UDS environment mostly maintains a constant low speed. In this flight state, the specific force measured by IMU is the projection of gravity in the IMU coordinate system. The α_{roll^r} and $\alpha_{\mathrm{pitch}^r}$ of UAVs can be corrected by using this specific force as a gravity constraint. The projection of the specific force component $\boldsymbol{f}_I$ measured by IMU in the body coordinate system can be expressed as $\boldsymbol{f}_B = \boldsymbol{R}_I^B \boldsymbol{f}_I$, where $\boldsymbol{R}_I^B$ can be obtained in advance by external parameter calibration. $\boldsymbol{g}_W$ is the gravity under the world coordinate system, and its projection under the body coordinate system can be expressed as $\boldsymbol{g}_B = \boldsymbol{R}_W^B \boldsymbol{g}_W$, where $\boldsymbol{R}_W^B$ can be obtained from the visual odometry.

When the UAV is flying at a constant speed, $\boldsymbol{f}_B$ and $\boldsymbol{g}_B$ should exhibit consistency. However, due to the accumulated error in the recursive process, the actual $\boldsymbol{f}_B$ does not align with $\boldsymbol{g}_B$. Consequently, a gravity constraint residual $\boldsymbol{\rho}$ is formulated to represent the error

$$\boldsymbol{\rho} = \arcsin\left(\frac{\boldsymbol{g}_B \times \boldsymbol{f}_B}{\| \boldsymbol{g}_B \| \cdot \| \boldsymbol{f}_B \|}\right) \tag{13}$$

In this paper, the gyro data is corrected by using $\boldsymbol{\rho}$, thus $\boldsymbol{R}_W^B$ given by VIO is corrected. The corrected gyro data $\boldsymbol{\omega}$ by $\boldsymbol{\rho}$ is represented as

$$\hat{\boldsymbol{\omega}} = \boldsymbol{\omega} + d \cdot \boldsymbol{\rho} \tag{14}$$

where d determines the weight of gravity constraint in gyro correction. When the UAV is in a steady-state or low maneuverability mode, the gravity constraint is trusted and integrated with the IMU to eliminate accumulated errors.

In this paper, the residual error is constructed based on the gravity constraint, and the gyro data is modified by complementary filtering. The accuracy of $\boldsymbol{R}_W^B$ provided by VIO is enhanced, significantly restraining the roll and pitch divergence of the UAV.

3 Experiment and Analysis

3.1 Experimental platform and environment

In this paper, a self-built experimental platform is used to collect the required test data in the corridor. The experimental platform and environment are respectively shown in Fig.5.

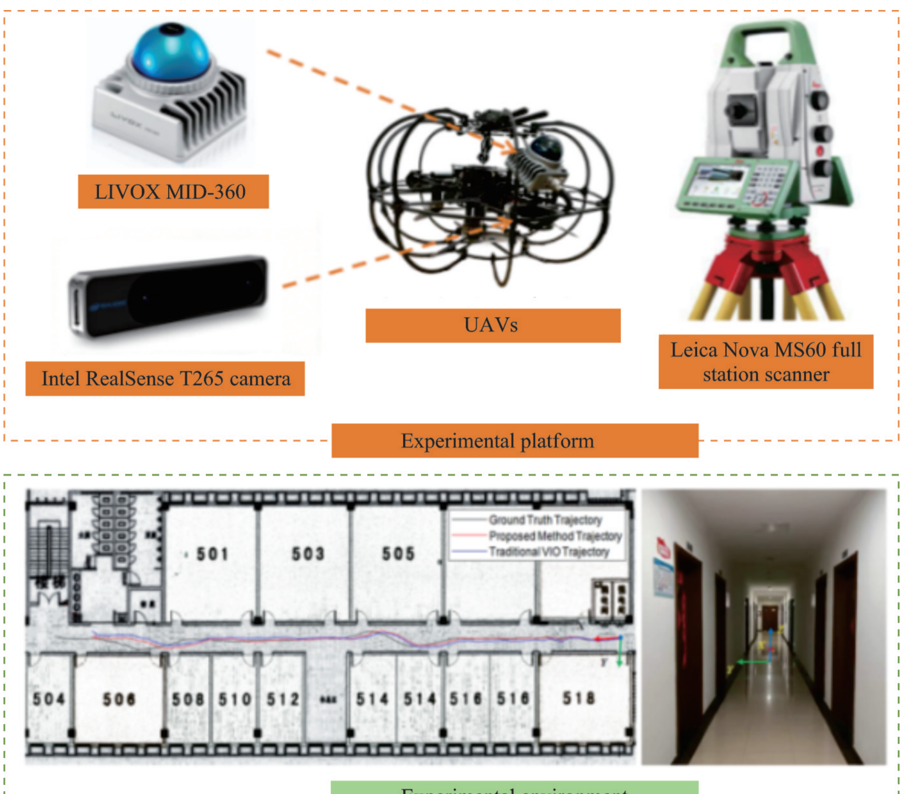


Fig.5 Experimental platform and environment

The equipment which is carried on the experimental platform includes LIVOX MID-360 LiDAR and Intel RealSense T265 camera. In Fig.5, the black line represents the ground truth trajectory, the red line the trajectory drawn by our proposed algorithm, and the blue line the trajectory drawn by traditional VIO. The proposed vision-inertial odometry algorithm based on structural information assistance is verified on the robot operating system (ROS). The computer utilized in this paper features a

3.3-GHz advanced micro device (AMD) eight-core CPU and 16 GB memory. Due to its high precision and robust visual odometry capabilities, the Intel RealSense T265 camera has been widely used in the UAV industry. Thus, in this experiment, the odometer information which is output by the Intel RealSense T265 camera is chosen to represent traditional visual odometry and is compared with the algorithm proposed in this paper.

For position reference, the experiment utilizes the Leica Nova MS60 full station scanner. The Leica Nova MS60 MultiStation is used to provide the ground-truth position measurements. In prism measurement mode, its distance measurement accuracy is specified as $\pm(1+0.001\ 5D)$ mm, where D is the measured distance in meters. As the total station scanner couldn't provide attitude reference, the LIVOX MID-360 LiDAR is employed, and the attitude angle calculated by FAST-LIO[29] serves as the attitude reference. The drift error is less than 0.05% of the total distance length of the carrier movement.

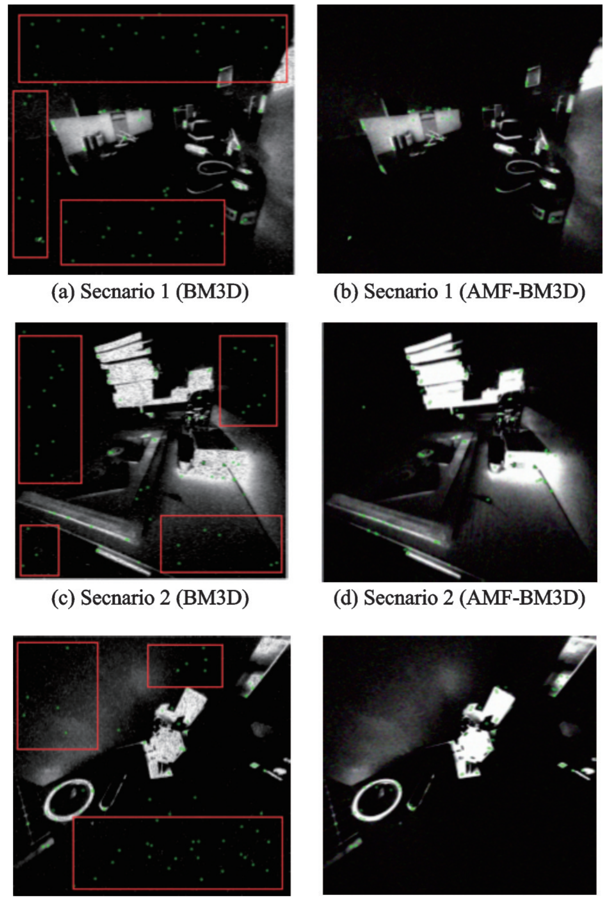
3.2 Analysis of image denoising

We employ the accelerated-KAZE (AKAZE) feature extraction algorithm, setting the number of feature points to 100. Feature points are sorted in descending order by response strength and selected preferentially, so that the quality of the denoised images can be directly reflected by the number and proportion of valid feature points. Three different scenes are selected for experiments, where denoising is performed using both BM3D and AMF-BM3D algorithms for comparison and analysis. The processing results are shown in Table 1 and Fig.6.

To further quantify the performance improvement, the ratio of valid feature points to the total number of feature points is calculated for three typical scenes. As shown in Table 1, AMF-BM3D achieves an average increase of approximate 38% in valid feature point ratio compared to traditional BM3D, fully demonstrating the superiority of the proposed method in denoising and feature extraction of weakly lit fisheye images.

Table 1 Ratio of effective feature points %

Algorithm	Scenario 1	Scenario 2	Scenario 3	Average
BM3D	50	57	44	50.33
AMF-BM3D	89	91	85	88.33



(a) Secnario 1 (BM3D) (b) Secnario 1 (AMF-BM3D)

(c) Secnario 2 (BM3D) (d) Secnario 2 (AMF-BM3D)

(e) Secnario 3 (BM3D) (d) Secnario 3 (AMF-BM3D)

Fig.6 Comparison of AMF-BM3D and BM3D algorithm effects

The items within the red box represent noise. Images denoised by BM3D still contain residual noise, resulting in sparse feature point distribution and more mismatches. In contrast, AMF-BM3D effectively suppresses noise, significantly increasing the number of feature points, which are more evenly and densely distributed. Feature extraction in key structural regions is more accurate and complete. Therefore, in this paper, AMF-BM3D is used to denoise the image to improve the feature point extraction effect, and the positioning effect is further verified on this basis.

3.3 Analysis of attitude and positioning accuracy

Limited by the length of the UDS environment, the total trajectory length for this experiment is 46 m. The coordinate system in the experimental environment image is the world coordinate constructed based on the characteristics of the environment. The results of attitude and positioning accuracy are shown in Tables 2, 3 and Fig.7. The root mean square error (RMSE) is employed to quantitatively evaluate the attitude and positioning errors.

Table 2 Comparison of attitude RMSE and drift

Algorithm	RMSE/(°)			Drift/%		
	Roll	Pitch	Yaw	Roll	Pitch	Yaw
Proposed method	0.88	0.22	1.67	0.28	0.29	0.29
Traditional VIO	1.42	0.55	2.73	0.46	0.71	0.48

Table 3 Comparison of position RMSE and drift

Algorithm	RMSE/m		Drift/%	
	Y axis	*Z* axis	*Y* axis	*Z* axis
Proposed method	0.03	0.03	0.62	6.69
Traditional VIO	0.12	0.23	2.78	49.26

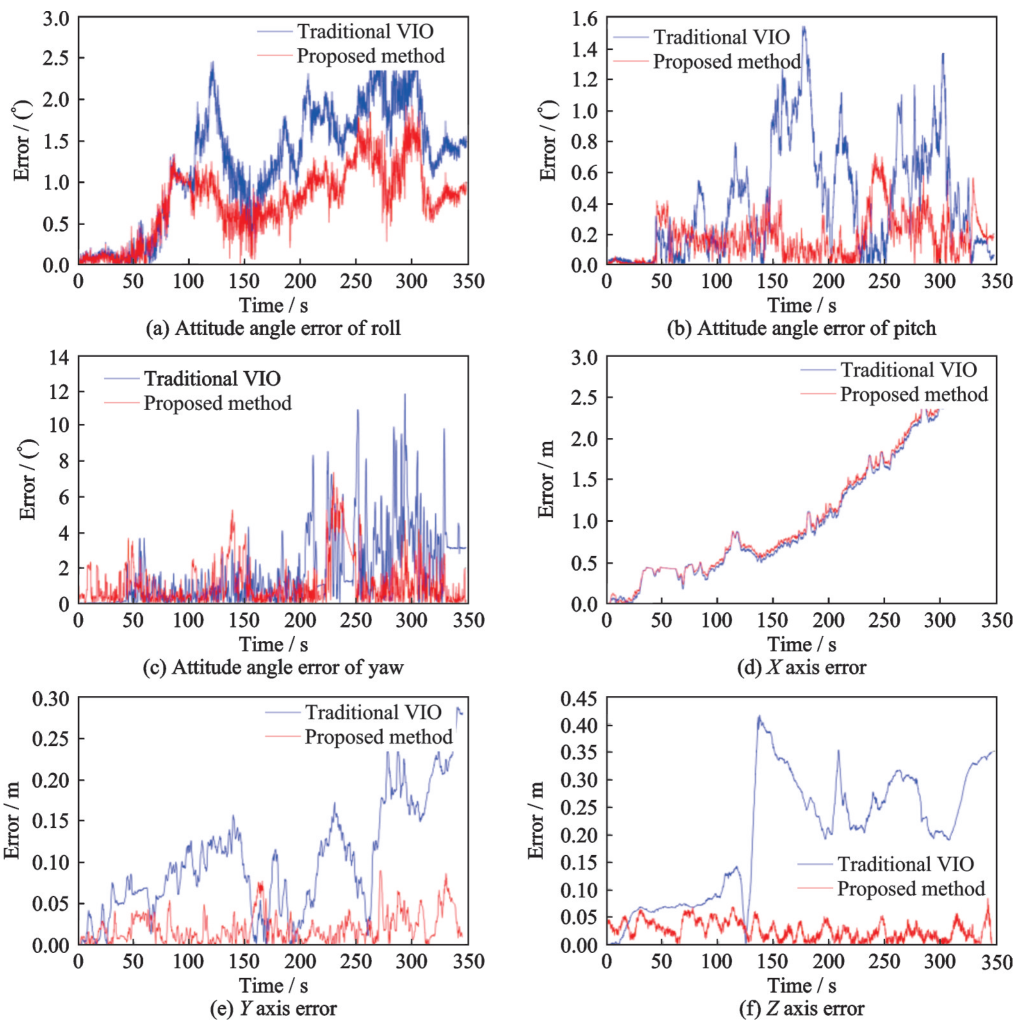


Fig.7 Attitude and position error curves

The following conclusions are drawn from Tables 2, 3 and Fig.7.

(1) Analysis of the data reveals that, in comparison to traditional VIO, the proposed algorithm

achieves an increased accuracy of roll, pitch and yaw in the UDS environment by 39.13%, 59.15% and 39.58%, respectively.

(2) The proposed algorithm optimizes the attitude of the UAV, leading to significant enhancements in Y axis and Z axis positioning. However, owing to the typical degraded characteristics of the UDS environment, the accuracy of motion estimation along the camera's optical axis is poor. Consequently, the positioning accuracy along the X axis exhibits a similarity to that of traditional VIO algorithms, characterized by notable errors.

Therefore, in this experiment, only the RMSE and drift of the Y axis and Z axis are compared. Analysis of the data indicates a substantial reduction in positioning error along the Y axis and Z axis compared to the traditional VIO algorithm, achieving an order of magnitude improvement in UDS environment. The proposed algorithm enhances positioning accuracy by 77.70% and 86.42%.

According to the experimental results, it can be observed that in UDS environment, traditional VIO methods suffer from rapidly diverging positioning errors along the Y axis and Z axis during long-distance movements. This issue arises due to the lack of relative attitude constraints. In contrast, the proposed method significantly improves attitude estimation in UDS environment, thereby enhancing overall positioning accuracy.

3.4 Real-time and robustness analysis

The core objective of this study is to achieve efficient and robust attitude estimation for aircraft. At the methodological level, we first make significant improvements to traditional vanishing point detection methods. By combining the MW assumption with gravity vector constraints, the computational process is significantly optimized.

At the algorithmic level, we improve the traditional vanishing point detection method. By combining the MW assumption with gravity vector constraints, we effectively narrow the search space for structural features. Experimental results indicate that this optimization significantly reduces computational redundancy. The processing frame rate of the attitude estimation module increases from 562.35 ms using the traditional exhaustive search strategy to 279.25 ms with the proposed gravity constraints. This data confirms that the algorithmic efficiency is improved by approximatel 50.3%, accelerating the visual module while maintaining estimation accuracy.

Furthermore, considering that structural features in UDS environments do not change abruptly between adjacent frames, it is unnecessary to perform heavy vanishing point calculations for every image frame. Therefore, we adopt an intermittent measurement to update strategy. The computationally intensive structural feature extraction is performed only on selected keyframes, while the high-frequency IMU integration maintains the pose estimation in the intervals. This strategy effectively balances the computational load and the estimation frequency.

To address the bottleneck issue where V-SLAM systems are prone to failure in extremely low-light or textureless environments, we innovatively introduce an IMU-based filtering algorithm for pose estimation. This method effectively fuses temporal data from the gyroscope and accelerometer, and through complementary filtering or a Kalman filter framework, can maintain stable pose output even when visual information is missing. As shown in Fig.8, in the simulated dark scenario, the traditional visual SLAM system (red line) exhibits significant divergence due to feature tracking failure, yielding mean errors of 7.23° in roll and 7.44° in pitch. In contrast, the proposed IMU-based complementary filtering method (blue line) effectively suppresses drift, restricting the mean error to 1.37° for roll and 1.89° for pitch. These metrics correspond to accuracy improvements of 81.05% and 74.59%, definitively demonstrating that the system can maintain high-precision pose estimation even when visual information is severely degraded.

After full-process system testing, the visual processing thread achieves an average frame rate of 12.28 Hz on standard test sequences. It is important to note that while the visual module operates at this rate, the final state estimation output of the system

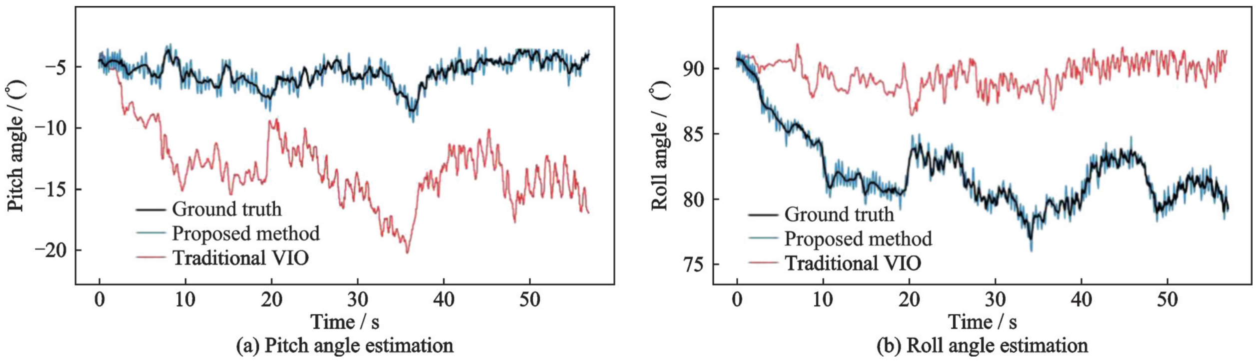


Fig.8 Attitude estimation in dark environment

is synchronized with the high-frequency IMU. By fusing the low-latency IMU propagation with the visual corrections, the system provides real-time pose feedback aligned with the IMU frequency, satisfying the high-bandwidth control requirements of UAVs in complex, dynamic environments.

4 Conclusions

This paper proposes a visual odometry method that integrates relative attitude information for UDS environment with extended longitudinal distance, narrow surroundings and a single extension direction. We estimate the relative heading angle using the vanishing point method and roll and pitch angles through complementary filtering. Through experiments conducted in UDS environment and compared with traditional VIO algorithms, the proposed method significantly improves the accuracy of attitude and position estimation in non-degraded directions of UDS environment by integrating relative attitude information, thus enhancing the flight safety of UAVs in narrow UDS environment.

References

[1] PU H Z, ZHEN Z Y, XIA M, et al. Unmanned aerial system and flight sontrol: A review[J]. Transactions of Nanjing University of Aeronautics & Astronautics, 2015, 32(1): 1-8.

[2] LI Z X, HE Q H, ZHAO H, et al. Doublem-net: Multi-scale spatial pyramid pooling-fast and multi-path adaptive feature pyramid network for UAV detection[J]. International Journal of Machine Learning and Cybernetics, 2024, 15(12): 5781-5805.

[3] LI Zeyu, WU Xian, LIU Huailong, et al. Review on vision/LiDAR SLAM for unmanned aerial vehicle swarms[J]. Navigation Positioning and Timing, 2025, 12(4): 15-26.(in Chinese)

[4] LI Z H, ZHANG Y B. Constrained ESKF for UAV positioning in indoor corridor environment based on IMU and WiFi[J]. Sensors, 2022, 22(1): 391.

[5] JAFAR F A, SUZUKI Y, TATENO Y, et al. An environmental visual features based navigation for mobile robot in a corridor environment[C]//Proceedings of 2010 IEEE International Conference on Robotics and Biomimetics. Tianjin, China: IEEE, 2010: 1612-1617.

[6] KUMAR G A, PATIL A K, PATIL R, et al. A LiDAR and IMU integrated indoor navigation system for UAVs and its application in real-time pipeline classification[J]. Sensors, 2017, 17(6): 1268.

[7] CADENA C, CARLONE L, CARRILLO H, et al. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age[J]. IEEE Transactions on Robotics, 2016, 32(6): 1309-1332.

[8] CHEN K Q, XIAO J H, LIU J L, et al. Semantic visual simultaneous localization and mapping: A survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2025, 26(6): 7426-7449.

[9] QIN T, LI P L, SHEN S J. VINS-mono: A robust and versatile monocular visual-inertial state estimator[J]. IEEE Transactions on Robotics, 2018, 34(4): 1004-1020.

[10] FORSTER C, PIZZOLI M, SCARAMUZZA D. SVO: Fast semi-direct monocular visual odometry[C]//Proceedings of 2014 IEEE International Conference on Robotics and Automation (ICRA). Hong Kong, China: IEEE, 2014: 15-22.

[11] LIU X D, GAO Z, CHENG H M, et al. Learning-based low light image enhancement for visual odometry[C]//Proceedings of 2020 IEEE 16th International Conference on Control & Automation (ICCA). Singapore: IEEE, 2020: 1143-1148.

[12] WANG Mengqi, LIAN Zengzeng, TIAN Yalin, et

al. Research on visual and inertial navigation fusion algorithms for indoor low-light environments[J]. Navigation Positioning and Timing, 2024, 11(5): 25-35. (in Chinese)

[13] DABOV K, FOI A, KATKOVNIK V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering[J]. IEEE Transactions on Image Processing, 2007, 16(8): 2080-2095.

[14] CHEN Y J, POCK T. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1256-1272.

[15] MOURIKIS A I, ROUMELIOTIS S I. A multi-state constraint Kalman filter for vision-aided inertial navigation[C]//Proceedings of 2007 IEEE International Conference on Robotics and Automation. Rome, Italy: IEEE, 2007: 3565-3572.

[16] CAMPOS C, ELVIRA R, RODRIGUEZ J J G, et al. ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.

[17] LI H, SAVKIN A V. An optical flow based tunnel navigation algorithm for a flying robot[C]//Proceedings of 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO). Kuala Lumpur, Malaysia: IEEE, 2018: 1767-1770.

[18] ELMOKADEM T, SAVKIN A V. A method for autonomous collision-free navigation of a quadrotor UAV in unknown tunnel-like environments[J]. Robotica, 2022, 40(4): 835-861.

[19] WANG L Q, XU H, ZHANG Y C, et al. Neither fast nor slow: How to fly through narrow tunnels[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 5489-5496.

[20] GAO B, WANG D J, LIAN B W, et al. LOVINS: Lightweight omnidirectional visual-inertial navigation system[C]//Proceedings of 2021 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). Xi' an, China: IEEE, 2021: 1-6.

[21] WEN S H, ZHAO Y F, LIU X, et al. Hybrid semi-dense 3D semantic-topological mapping from stereo visual-inertial odometry SLAM with loop closure detection[C]//Proceedings of IEEE Transactions on Vehicular Technology. [S.l.]: IEEE, 2020: 16057-16066.

[22] WANG Y Y, MA X R, WANG J, et al. Robust AUV visual loop-closure detection based on variational autoencoder network[J]. IEEE Transactions on Industrial Informatics, 2022, 18(12): 8829-8838.

[23] SUN H, JIN Y Q, FU M L, et al. A multisensor-based tightly coupled integrated navigation system[C]//Proceedings of 2022 5th International Symposium on Autonomous Systems (ISAS). Hangzhou, China: IEEE, 2022: 1-6.

[24] CLARK R, WANG S, WEN H K, et al. VINet: Visual-inertial odometry as a sequence-to-sequence learning problem[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017, 31(1): 3995-4001.

[25] WANG S, CLARK R, WEN H K, et al. DeepVO: Towards end-to-end visual odometry with deep recurrent convolutional neural networks[C]//Proceedings of 2017 IEEE International Conference on Robotics and Automation (ICRA). Singapore: IEEE, 2017: 2043-2050.

[26] LI R H, WANG S, LONG Z Q, et al. UnDeepVO: Monocular visual odometry through unsupervised deep learning[C]//Proceedings of 2018 IEEE International Conference on Robotics and Automation (ICRA). Brisbane, Australia: IEEE, 2018: 7286-7291.

[27] NEYESTANI A, PICARIELLO F, AHMED I, et al. From pixels to precision: A survey of monocular visual odometry in digital twin applications[J]. Sensors, 2024, 24(4): 1274.

[28] LU X H, YAO J, LI H A, et al. 2-line exhaustive searching for real-time vanishing point estimation in Manhattan world[C]//Proceedings of 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). Santa Rosa, USA: IEEE, 2017: 345-353.

[29] XU W, ZHANG F. FAST-LIO: A fast, robust LiDAR-inertial odometry package by tightly-coupled iterated Kalman filter[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 3317-3324.

Acknowledgements This work was supported by the Basic Research Program of Jiangsu Province (No. BK20250175) and the National Natural Science Foundation of China (No. 62273178).

Authors

The first author Mr. **XIE Shenghong** is currently pursuing the Ph.D. degree at College of Automation Engineering, Nanjing University of Aeronautics and Astronautics (NUAA), China. He received his bachelor degree in automation from NUAA in 2023. His research interest includes spanning the field of UAV visual navigation.

The corresponding author Prof. **LYU Pin** received the B.S. degree in electrical engineering and automation and the Ph.D. degree in navigation, guidance and control from NUAA, Nanjing, China, in 2008 and 2015, respectively. From 2015 to 2016, he conducted postdoctoral research at the Aerospace Research Institute of the University of Toronto. Since 2017, he has been working at the College of Automation Engineering, NUAA. His research interests include UAV navigation, autonomous navigation, visual navigation, combined navigation, and low-altitude aircraft navigation.

Author contributions Mr. **XIE Shenghong** was responsible for research design, method development, experimental analysis, and paper writing. Mr. **WEN Yebei** provided experimental design and technical support. Mr. **WANG Hongfeng** was responsible for paper proofreading and equation checking. Mr. **HUANG Zheng** conducted the literature review and wrote the research background. Mr. **JI Tianyang** assisted with experimental verification and data analysis. Mr. **YUAN Cheng** provided suggestions for manuscript revision. Prof. **LYU Pin** provided guidance on research design and paper writing. All authors commented on the manuscript draft and approved the submission.

Competing interests The authors declare no competing inrests.

(Production Editor: WANG Jie)

单向结构信息辅助的无人机视觉惯性里程计

谢盛弘[1]，温烨贝[1]，王洪枫[2]，黄 郑[1]，季田阳[1]，袁 诚[1]，吕 品[1]

（1.南京航空航天大学自动化学院，南京 210016，中国；2.北京自动化控制设备研究所，北京 100074，中国）

摘要：针对地下综合管廊、隧道及长距离走廊等单向结构(Unidirectional structural，UDS)环境中光照不足、纹理特征稀疏等因素导致视觉惯性里程计特征提取不稳定、姿态误差持续累积和无人机(Unmanned aerial vehicles，UAVs)定位精度下降的问题，提出一种单向结构信息辅助的无人机视觉惯性里程计方法。该方法充分利用UDS环境沿纵向延伸、侧壁近似平行以及结构方向长期稳定的特点，将环境结构信息作为无人机姿态估计的外部约束，实现对传统视觉惯性里程计累计误差的持续修正。该方法通过采用自适应中值滤波与BM3D(Adaptive median filtering and BM3D, AMF-BM3D)去噪方法来提高特征提取精度，并引入一种惯性容错消失点技术，以实现高精度、无累积误差的无人机偏航角估计；同时利用重力约束来消除滚转角与俯仰角中的累积误差。最后，采用扩展卡尔曼滤波器(Extended Kalman filter, EKF)对传统视觉—惯性里程计(Visual-inertial odometry，VIO)获取的位姿数据进行融合，从而获得精细化的位姿估计结果。在UDS环境中的实验验证表明，与传统VIO方法相比，所提出的算法显著提升了UDS环境下的位姿估计精度及无人机飞行安全性。

关键词：结构信息；视觉惯性里程计；扩展卡尔曼滤波

研究亮点：

1. 提出面向暗光鱼眼图像混合噪声的AMF-BM3D预处理方法，通过椒盐噪声检测掩膜与自适应中值滤波、BM3D协同处理，实现对椒盐噪声和高斯噪声的针对性抑制，增强了弱光UDS环境下视觉前端的特征提取稳定性。

2. 提出基于惯性容错消失点法的相对偏航角估计算法。该算法在UDS环境下依据曼哈顿假设对相对偏航角进行估计，并利用短期惯性递归信息来维持偏航角在突发环境变化中的稳定性。该算法增强了相对偏航角估计的鲁棒性，有效缓解了随时间推移而产生的误差累积问题。

3. 构建重力约束姿态校正与结构信息辅助视觉惯性融合方法，根据无人机在UDS环境中的低速、小机动运动特性，利用重力约束和互补滤波持续修正滚转角与俯仰角，并结合EKF完成结构姿态信息与视觉惯性里程计融合，使滚转、俯仰和航向估计精度分别提高39.13%、59.15%和39.58%，同时显著降低非退化方向的位置漂移。