Research Article

Applying SLAM Algorithm Based on Nonlinear Optimized Monocular Vision and IMU in the Positioning Method of Power Inspection Robot in Complex Environment

Can Wang, Zhibin Li, Yefei Kang, and Yingzheng Li

School of Automation Engineering, Shanghai University of Electric Power, Shanghai 200090, China

Correspondence should be addressed to Can Wang; can_wang@mail.shiep.edu.cn

Received 24 May 2022; Revised 1 September 2022; Accepted 5 September 2022; Published 19 September 2022

Academic Editor: Xos M. Pardo

Copyright © 2022 Can Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Under China’s Intelligent Electric Power Grid (IEPG), the research on IEPG inspection mode is of great significance. This work aims to improve the positioning and navigation performance of IEPG inspection robots in a complex environment. First, it reviews the monocular camera projection and the Inertial Measurement Unit (IMU) models. It also discusses the tight-coupling monocular Vision Inertial Navigation System (VINS) and the initialization theory of the Simultaneous Localization and Mapping (SLAM) system. Nonlinear optimization for SLAM by the Gauss–Newton Method (GNM) is established. Accordingly, this work proposes the SLAM system based on tight-coupling monocular VINS. The EuRoC dataset data sequence commonly used in visual-inertial algorithm testing in IEPG is used for simulation testing. The proposed SLAM system’s attitude and position estimation errors are analyzed on different datasets. The results show that the errors of roll, pitch, and yaw angle are acceptable. The errors of the X, Y, and Z axes are within 40 cm, meeting the positioning requirements of an Unmanned Aerial Vehicle (UAV). Meanwhile, the Root Mean Square Error (RMSE) evaluates the improvement of positioning accuracy by loop detection. The results testify that loop detection can reduce the RMSE and improve positioning accuracy. The attitude estimation tests the angle changes of pitch, roll, and yaw angles with time under a single rotation condition. The estimated value of the proposed SLAM algorithm is compared with the real value through Absolute Trajectory Error (ATE). The results show that the real value and the estimated value of attitude error can coincide well. Thus, the proposed SLAM algorithm is effective for positioning and navigation. ATE can also be controlled within ±2.5°, satisfying the requirements of navigation and positioning accuracy. The proposed SLAM system based on tight-coupling monocular VINS presents excellent positioning and navigation accuracy for the IEPG inspection robot. The finding has a significant reference value in the later research of IEPG inspection robots.

1. Introduction

With the rapid development of China’s Intelligent Electric Power Grid (IEPG) system and the proposal of the concept of a smart grid, the IEPG inspection mode has attracted extensive attention. People have favored the IEPG inspection robot to ensure the safe operation of the power system [1]. In particular, the IEPG inspection robot can make up for the low efficiency of manual inspection. It features solid environment robustness and low safety risk and improves power grid operation reliability. The IEPG inspection robot realizes inspection by leaning upon its independent navigation and positioning functions [2]. Therefore, the accurate positioning, strong robustness, and strong environmental adaptability of the IEPG inspection robot are of great significance to the power system [3]. Improving the IEPG inspection robot’s these features has become the focus of relevant research. Simultaneous Localization and Mapping (SLAM) is an important technology to realize intelligent robots’ self-active navigation and positioning in unfamiliar environments. It is an essential part of environmental perception [4]. SLAM uses its vision sensor to collect image sequences and obtain rich external scene information similar to that observed by human eyes. With the continuous innovation of vision sensors, the vision-based SLAM method has become a research hotspot [5].
However, the vision-based SLAM system has some disadvantages, such as lack of texture information, greatly affected by external light, inability to work normally in high dynamic light or dark environments, and fast speed-triggered motion blur and motion estimation failure [6]. Affected by the detection and tracking of visual front-end feature points, the visual SLAM system relies heavily on the quality of the image. A single-scene environment without rich features will generate a blurred image when the carrier moves too fast. Sometimes, two frames will be overlapped substantially. As a result, the feature matching fails, and the depth information of the pixel cannot be obtained, resulting in the lack of absolute scale, scale drift, and other problems [7]. These problems can be solved by fusing the vision sensor with other sensor information. For example, there is a complementary relationship between the IMU (Inertial Measurement Unit) and the monocular camera. The absolute scale and gravity direction during the carrier movement can be obtained by fusing IMU data. In particular, IMU can provide relatively accurate attitude estimation in a short time in the case of rapid carrier movement [8]. IMU can help the monocular camera determine the scale information and estimate the short-time attitude when the camera is blocked. Thus, it improves the positioning accuracy of the navigation system [9]. There are still some problems in fusing IMU with the monocular camera, such as over-reliance on the hardware design, off-line calibration of sensor parameters by the toolbox, complicated initialization, and a large amount of calculation, low precision, and low robustness on low-cost sensors [10]. Therefore, it is crucial to further optimize the monocular camera and IMU system. There are many kinds of research on IEPG inspection robots. Chen et al. (2020) took Daxing International Airport as an example. They studied the IEPG inspection robot and proposed the implementation scheme of installing the IEPG inspection robot in the switching station of Daxing International Airport [11]. Zhong et al. (2021) used Genetic Algorithm (GA) to determine the running image of the IEPG inspection robot and calculate the offset angle by analyzing the scale characteristics. They aimed to study the optimization method of job scheduling [12]. Zhang et al. (2022) designed an IEPG inspection system for power line Unmanned Aerial Vehicle (UAV) by using robot trajectory tracking technology [13]. In terms of IMU localization research, Poulse and Han (2019) proposed a smartphone camera-based localization system and a hybrid localization system, to minimize the localization error in smartphone cameras and IMU-based localization systems, and analyzed the performance of the system [14]. Poulse et al. (2019) proposed a position estimation algorithm that estimates pedestrian heading position using a combination of features from the accelerometer, magnetometer, and gyroscope data from IMU sensors. Pitch and roll values and gyroscope sensor values are estimated based on the combined function of the accelerometer, magnetometer, and gyroscope data from the IMU position sensor and fusion of the accelerometer [15]. On account of the SLAM, Jia et al. (2022) studied and proposed the mapping and positioning methods of inspection robots, respectively and further proposed an inspection control method based on iterative learning control (Jia et al., 2022) [16].

Firstly, the monocular camera projection model is detailed. Secondly, the IMU sensor measurement model and preintegration are described in detail. Finally, the image data preprocessing method is elaborated, and the related theory of nonlinear optimization of the SLAM system is introduced. On the basis of the above-given theory, the nonlinear optimization process of SLAM based on the Gauss–Newton Method (GNM) is formed, and the structure of the SLAM algorithm integrating vision and IMU under nonlinear optimization is constructed. The resulting test of the tight-coupling VINS algorithm is studied, the influence of loop detection on positioning accuracy is analyzed, and the test results of attitude estimation under rotation conditions are explored in detail. The designed algorithm is compared with the Root Mean Square Error (RMSE) obtained by other algorithms. The main contribution of the research is to improve and optimize the original SLAM system and establish a nonlinear optimization process based on the GNM to solve SLAM. And, the structure of the SLAM algorithm integrating vision and IMU under nonlinear optimization is constructed. The design aims to improve the estimation accuracy, attitude positioning accuracy, computational speed, and rapid deployment capability of the monocular VINS of the power inspection robot, and it is expected to provide a technical reference for the improvement of the working performance of the power inspection robot.

2. Research Theory and Method Design

The overall frame structure is designed, and the overall structure is presented in Figure 1.

In Figure 1, on the basis of the monocular camera projection model, IMU sensor measurement model and preintegration, image data preprocessing method, and the related theory of nonlinear optimization of the SLAM system, the GNM-based nonlinear optimization process of SLAM is constructed, and the SLAM algorithm structure of vision and IMU fusion under nonlinear optimization is further formed. Then, the experimental test and data processing are carried out, and the detailed analysis is performed from four aspects. They are the resulting test of the tight-coupling VINS algorithm, the influence of the loop detection on the positioning accuracy, the test of attitude estimation under the rotation condition, and the error analysis of different algorithms, respectively.

2.1. The Monocular Camera Projection Model. There are many kinds of vision sensors: monocular camera, binocular camera, depth camera, and panoramic camera. The latter three have significant advantages and disadvantages. The monocular camera only needs one camera, which is low cost, small volume, and high compatibility with application scenes [17]. The monocular camera and IMU are the two most important sensors in the VINS. IMU is responsible for measuring the inertial information of the carrier, establishing the connection between visual observation and the
inertial coordinate system, and the camera samples the environmental information [18]. A monocular camera connects Three-Dimensional (3D) (spatial) coordinate points with Two-Dimensional (2D) (plane) coordinate points through projection and pixelation. Monocular Camera Projection Mapping mainly involves a world coordinate system, camera coordinate system, image coordinate system, and pixel coordinate system [19]. The world coordinate system is the coordinate system fixed on the Earth’s surface. The camera coordinate system refers to the coordinate system fixed on the camera and moving with the movement of the camera. The Pinhole Camera Model (PCM) is widely used in the coordinate system. In this work, the PCM is used to solve the internal parameters of the camera [20]. Figure 2 shows the projection principle of PCM.

In Figure 2, O means the camera’s optical center, which is the pinhole of the PCM. \( O - X - Y - Z \) represents the camera coordinate system. A point \( Q \) in space, projected through hole \( O \), falls on \( O' - X' - Y' - Z' \) to form the imaging point \( Q' \). Then, the coordinate of point \( Q \) is \( [X; Y; Z]^T \). The coordinate of \( Q' \) is \( [X'; Y'; Z']^T \). The focal length is denoted by \( f \). According to the triangle principle, the following equation holds:

\[
\frac{Z}{f} = \frac{X}{X'} = \frac{Y}{Y'}
\]  

(1)

Finally, the values of \( X' \) and \( Y' \) are obtained:
\[ X' = f \frac{X}{Z}, \quad Y' = f \frac{Y}{Z} \]  
(2)

Equation (2) describes the relationship between spatial point \( Q \) and imaging. It is converted to a pixel plane, defines the pixel coordinates, and scales them on the \( u \)-axis and \( v \)-axis, respectively, \( \alpha \) times and \( \beta \) times. Then, the origin translation amount is \([c_x, c_y]^T\). The relationship between point \( Q \) coordinates and pixel coordinates is illustrated in the following equation:

\[
\begin{cases}
  u = \alpha X' + c_x \\
  v = \beta Y' + c_y.
\end{cases}
\]  
(3)

Here, \( \alpha f = f_x \), and \( \beta f = f_y \). Then, the following equation can be obtained:

\[
\begin{cases}
  u = f_x \frac{X}{Z} + c_x \\
  v = f_y \frac{Y}{Z} + c_y.
\end{cases}
\]  
(4)

Here, \( f \) is calculated by \( \text{mm} \). \( \alpha \) and \( \beta \) are counted by pixel/mm and are converted to matrix form. Then, the following equation is obtained:

\[
\begin{pmatrix}
  u \\
  v \\
  1
\end{pmatrix} = \frac{1}{Z} \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \triangleq \frac{1}{Z} \text{KP}.  
\]  
(5)

2.2. IMU Sensor Measurement Model and Preintegration. The first task of the visual SLAM front-end is to extract and track feature points from image frames based on the model imaged by the camera (Danping et al., 2019) [21]. The main components of the IMU are accelerometers and gyroscopes, which mainly measure the acceleration and angular velocity of the moving carrier and estimate the pose of the carrier (Ashry et al., 2020) [22]. Considering the bias and noise in the measurement process, the IMU ignores the influence of the Earth’s rotation, and the measurement output of the 3D space-based M-IMU can be simply modeled, which is defined as follows:

\[ B\omega_{WB}(t) = B\omega_{WB}(t) + b^\theta(t) + \eta^\theta(t), \]  
(9)

\[ B\alpha(t) = R_{BW}(t)(Wa(t) - Wg) + ba(t) + \eta^\alpha(t). \]  
(10)

\( B \) represents the IMU coordinate system; \( W \) refers to the world coordinate system, respectively; \( g \) and \( a \) are the gyroscope and accelerometer, respectively. \( b^\theta(t) \) and \( b^a(t) \) denote the zero bias of the two sensors. \( \eta^\theta(t) \) and \( \eta^\alpha(t) \) signify the measurement noise of two sensors. \( B\omega_{WB}(t) \) is the actual value of the angular velocity of the carrier system. \( R_{BW}(t) \) means the rotation matrix of the \( B \) system and \( W \) system. \( Wa(t) \) stands for the acceleration of the carrier in the world coordinate system. \( Wg \) signifies the gravitational acceleration of the carrier in the world coordinate system.

The noise of the accelerometer and gyroscope follows the Gaussian distribution. The specific expression is as follows:

\[
\begin{align*}
E(\eta^\alpha(t)) &= 0, \\
E(\eta^\theta(t)) &= 0, \\
E[\eta^\alpha(t_1)\eta^\alpha(t_2)] &= \sigma^\alpha_1\delta(t_1 - t_2) \\
E[\eta^\theta(t_1)\eta^\theta(t_2)] &= \sigma^\theta_1\delta(t_1 - t_2).
\end{align*}
\]  
(11)

\( \sigma^\alpha \) and \( \sigma^\theta \) represent the noise intensity of the accelerometer and gyroscope, and the random walk process of the two is modeled as Brownian motion, it can be obtained as follows:

\[ b^\alpha(t) = \eta^{\alpha_1} + b^\alpha(t) = \eta^{\alpha_2}. \]  
(12)

\( \eta^{\alpha_1} \) and \( \eta^{\alpha_2} \) are white Gaussian noise with zero mean. Through the above-given modeling, the relationship between two image frames \([t_j, t_{j+1}]\) can be obtained, and the specific expression is as follows:

\[
\begin{align*}
b^\theta(t_j) &= b^\theta(t_j) + \eta^{\theta_1} \\
b^\alpha(t_j) &= b^\alpha(t_j) + \eta^{\alpha_1}.
\end{align*}
\]  
(13)

\( \eta^{\theta_1} \) and \( \eta^{\alpha_1} \) indicate zero mean. The covariance is the white noise of the following equation:

\[
\sum = \Delta t_{ij}\text{cov}(\eta^\theta) + \sum = \Delta t_{ij}\text{cov}(\eta^\alpha). \]  
(14)

The acquisition frequency of IMU is 200 Hz, which is much higher than the frame rate of the image collected by the camera. Therefore, the IMU data must be preprocessed.
The IMU data between two adjacent keyframes can be combined into a composite term to constrain the two visual keyframes through IMU preintegration.

2.3. Image Data Preprocessing. Facing the large-scale scene of the power grid, the inspection robot tracks and extracts the feature points. This is mainly because the camera motion is estimated based on the extraction and matching of feature points, and the reprojection error is optimized, which has high robustness to scenes with large changes such as illumination, motion, and rotation. The preprocessing of the original image data collected by the camera can be divided into three modules. The first module tracks the image feature points. The second module packages the tracked feature points into point clouds and publishes them according to the given frequency. The last module builds the observation map model according to the published point clouds [23]. Feature point tracking is also called landmark point tracking and is a more characteristic pixel in a frame image. Figure 3 lays out the tracking process of feature points.

In Figure 3, the obtained new image is judged as whether it is the first frame image. A certain amount of feature information is evenly extracted from the image if it is the first frame image. The feature information is numbered. Conversely, if it is not the first frame image, the optical flow method is used for tracking to move out the lost tracking points and judge whether to publish them. The tracking process returns to the initial state if the points are not published. If the points are published, the external points will be removed, and new feature points will be added.

For tight-couple monocular VINS, IMU acquisition initialization provides system optimization parameters and initial state values, which is its key [24]. The variables mainly include gyro bias, scale factor, gravity, and velocity vector in the initialization process. Figure 4 showcases the whole process of IMU initialization.

Apparently, the first step for IMU initialization is the gyro offset estimation. When the camera continuously collects N keyframes, the rotation transformation matrix of adjacent image frames can be solved through monocular vision 3D reconstruction. The second step is the approximate estimation of scale and gravity acceleration, the zero bias estimation of the accelerometer, and the correction of scale and gravity acceleration. Lastly, the velocity is estimated.

Visual inertial joint initialization is the prerequisite to realizing the overall attitude control of the robot. The optimal attitude control of monocular VINS is estimated by combining IMU constraints and IMU-based ones. Figure 5 demonstrates the process of tight-coupling monocular VINS.

In Figure 5 demonstrates that features are extracted and tracked by a vision sensor. The inertial information is preintegrated and then fused to obtain the visual-inertial joint initialization. Furthermore, tight-coupling optimization is carried out to obtain the optimized attitude state. The sampling difference between the IMU and the camera can be eliminated, the IMU can provide scale information for monocular vision, and the visual pose can correct the drift problem of the IMU, which can reduce the probability of large errors.

In addition to the constraints of landmark observation data between continuous image frames, there are two other constraints introduced by IMU. For example, there are constraints on the speed, displacement, and rotation of IMU preintegration between continuous image frames. The random walk process constrains the zero deviation of IMU corresponding to two adjacent image frames [25]. According to the constraints, the factor graph can be constructed. Figure 6 describes the factor graph of VINS.

Figure 6 illustrates that the factor graph is usually composed of factors and variables. In Figure 5, the block is the state variable to be estimated, and the circular node denotes the constraint relationship between variables, including IMU constraint, visual constraint, and loop constraint. Purple nodes represent camera attitude, and brown nodes signify IMU measurements. Each factor in Figure 6 corresponds to an error term in the nonlinear Least Squares Problem (LSP).

2.4. Theories Related to Nonlinear Optimization of SLAM Systems. SLAM system is mainly used to solve the problem of positioning, navigation, and mapping of mobile robots in a complex environment [26]. It can estimate the motion state of the camera’s own degrees of freedom by means of multiview geometry and extended nonlinear optimization based on a series of image sequences with prior information in an unknown environment [27]. Figure 7 expresses the overall architecture of SLAM.

As exhibited in Figure 7, the SLAM system can be divided into visual odometry (VO), back-end optimization, Loop Closing, and drawing. It can complete navigation and positioning tasks in various environments. The basic task of monocular vision SLAM is to determine the position and attitude of the camera and the 3D coordinates of feature points in an unknown environment. In monocular vision SLAM, only noisy camera images can be obtained. Thus, global nonlinear optimization is needed. Integrating the IMU sensor into SLAM can further obtain carrier acceleration and angular velocity information. It helps estimate the position and attitude of the carrier and the coordinates of landmarks and realize the positioning of an intelligent machine. However, the information contains noise and needs nonlinear optimization. Loop detection plays a vital role in the vision SLAM system. Although the odometer can get the camera attitude map and the position information of feature points, the odometer only considers the correlation information of adjacent frames. It does not establish a consistent optimization trajectory globally. This results in the drift during the SLAM system running. In particular, the loop detection function can minimize the drift of the SLAM system and improve its accuracy.

The monocular VINS-based SLAM must be initialized before running [28]. The aim is to guide the system state and generate the initial state information (namely, the initial system parameter) [29]. The subsequent positioning and
mapping process is closely related to the system’s initial state information. Figure 8 draws the initialization flow of the nonlinear optimized SLAM system.

Figure 8 depicts that system initialization is divided into two processes. In the first process, the initial state information is estimated by visual initialization to obtain inaccurate state information. Then, the joint inertial navigation is initialized by IMU to improve the accuracy of the initialization state information. First, the visual sensor obtains the data information, and the keyframes are filtered according to certain conditions. The first keyframe is obtained as the current frame, and another keyframe is taken as the current frame to replace the first keyframe. Then, the first keyframe becomes the previous frame. Extracting the feature points can obtain the feature information of the two keyframes. Furthermore, triangulation calculation is performed to obtain the motion attitude information with the position feature. Next, the information map is established and continuously optimized, and finally, the system initialization information can be obtained.

2.5. Solving Nonlinear SLAM Optimization Process Based on GNM. The nonlinear SLAM optimization process is derived using the GNM. In visual SLAM, the sum of squares of the error term is generally represented by a cost function [30]. The specific expression is as follows:

$$F(X) = \frac{1}{2}e^T R^{-1} e.$$  \hspace{1cm} (16)

The error is expanded by first-order Taylor and is linearly approximated to obtain:

$$e(X + \Delta X) \approx \bar{e} + J\Delta X.$$  \hspace{1cm} (17)

$X \triangleq X_{n-1}$ is the state quantity at time $n - 1$ and is the column vector. $\bar{e} \triangleq e(X)$ denotes the error at time $n - 1$ and is the column vector. $J \triangleq \partial e/\partial X/X$ represents the first derivative of the error with respect to the state quantity. Substituting the above-given terms into (17) gets the following equation:

$$F(\bar{X} + \Delta X) \approx \frac{1}{2}e^T R^{-1} e + e^T R^{-1} \Delta X + \frac{1}{2}\Delta X^T J^T R^{-1} J \Delta X.$$  \hspace{1cm} (18)

The derivative of the increment $\Delta X$ is calculated so that it equals zero, resulting in the optimal solution for the increment:
Visual information
Feature information extraction and tracking

Pre integration data processing
Data correction

Feature information extraction

Visual inertia joint initialization

Visual inertial tight coupling optimization

Get posture

**Figure 5:** The process of tight-coupling monocular VINS.

**Figure 6:** Factor graph of VINS.
Back end: nonlinear optimization

Loop closing

Build map

Fore end: Visual Odometry

Sensor system

Figure 7: The overall framework of SLAM.

\[ \Delta X = \left( J^T R^{-1} \right)^{-1} J^T R^{-1} \hat{e}, \]  
\[ X = \tilde{X} + \Delta X. \]  

The final equation of increment \( \Delta X \) is also called the nonlinear least-squares equation, \( H = J^T R^{-1} J \) is defined, where \( H \) is the approximation of the Hessian matrix in Newton’s iterative method, which reduces the large computational consumption caused by Newton’s method for finding the Hessian matrix [31]. Then, \( \Delta X \) is used to calculate the iteration step size along the descending direction of the gradient until the visual reprojection error converges to the minimum. In the end, the optimal estimation is obtained for camera attitude and landmarks [32].

Construction of the proposed SLAM system based on tight-coupling monocular VINS under nonlinear optimization.

SLAM system can improve the positioning accuracy and robustness by fusing multisensors. In particular, fusing a simple and low-cost camera and IMU sensor kit is the best choice for multi-sensor fusion in theory and practical application [33]. Fusing IMU and monocular vision through nonlinear optimization is the framework with the best accuracy and robustness. It is essentially a tight-coupling monocular VINS. Figure 9 details the proposed SLAM system initialization, the system starts tracking, local map optimization, and closed-loop detection simultaneously. The tracking process is also called mileage. Its core part is to estimate the transformation relationship between two adjacent frames.

2.6. Experimental Test and Data Processing. The EuRoC dataset in IEPG is adopted, which can be used for simulation testing of visual algorithms and visual-inertial algorithms, and it is one of the most commonly used datasets for evaluating the positioning accuracy of visual SLAM algorithms. The experiment is divided into two parts. One is to use the hardware platform to test the error analysis of the algorithm and the analysis of the positioning accuracy. The other is to test the dynamic performance of the moving scene under the single rotation condition of the tight-coupling monocular VINS according to relevant literature research [34]. Table 1 shows the test environment of this experiment.

Absolute Trajectory Error (ATE) is usually used to evaluate the positioning accuracy of the algorithm. The specific calculation reads:

\[ \text{ATE}_{\text{rot}} = \left( \frac{1}{N} \sum_{i=0}^{N-1} \| \zeta (\Delta R_i) \| \right)^{1/2}, \]  
\[ \text{ATE}_{\text{pos}} = \left( \frac{1}{N} \sum_{i=0}^{N-1} \| \Delta \rho_i \| \right)^{1/2}. \]

In (18) and (19), \( \zeta (\Delta R_i) \) means converting the rotation matrix into an angle and using the angle to represent the attitude error. \( \Delta \rho \) stands for position error.

The RMSE is used to evaluate the improvement effect of loop detection on positioning accuracy. The expression is as follows:

\[ \text{RMSE} = \sqrt{\frac{N}{N-1} \sum_{i=0}^{N-1} \left( X_{\text{pi}} - X_{\text{Mi}} \right)^2}. \]

Xpi and XMi denote the estimated value and the actual value, respectively. N refers to the number of estimation points, and \( N > 1 \). According to the calculated RMSE, the smaller RMSE is, the more accurate the estimation is, and the better the estimation effect is.

The MH_06_difficult and MH_02_easy two dataset sequences contain sufficient rotation, translation, and highly representative illumination changes. Therefore, this work makes a quantitative analysis of the attitude error and position error of the positioning algorithm utilizing these two different data sequences. The proposed algorithm is implemented through the robot operating system.

3. Results and Discussion

3.1. Test and Analysis Results of Tight-Coupling VINS Algorithm. On account of the MH_06_difficult dataset sequence, the error of the proposed SLAM system based on tight-coupling monocular VINS is analyzed by quantifying the attitude error and position error. Figure 10 plots the
proposed algorithm’s state errors on the MH_06_difficult dataset sequence.

In Figure 10, for the attitude error, as the distance increases, the error fluctuation range of the heading angle becomes larger, and the error of the heading angle is larger than that of the pitch angle and roll angle. It is mainly because the pitch angle and roll angle are absolutely considerable and will not produce angular drift. Compared with the other two angles, the error of the roll angle and pitch angle is within ±2°. The error of the heading angle is larger, and the range is within ±3°. In Figure 9, for the position error, the fluctuation of the position error of the Y axis is the largest, followed by the X axis, and finally the Z axis. However, overall, the errors of the X, Y, and Z axes are all within 40 cm, which meets the positioning requirements of the UAV.

Subsequently, the proposed SLAM system by tight-coupling monocular VINS is evaluated on MH_02_easy dataset sequences by quantifying the attitude and position errors. Figure 11 reveals the results.

In Figure 11, for the MH_02_easy sequence, the attitude error of the heading angle is larger than that of the pitch and roll angles. The accuracy of the heading angle is compared with the other two angles, and the errors of the roll and pitch angles are within ±1°. The error of the heading angle is large, the range of the error angle is within ±3°, and the variation range is large. For the position error, the error of the Y-axis is the most obvious, followed by the X-axis, and finally the Z axis. However, on the whole, the three-axis errors of X, Y, and Z are all controlled within 300 mm, which can satisfy the positioning requirements of the UAV.

3.2. Influence of Loop Detection on Positioning Accuracy.
This section evaluates the proposed SLAM system based on tight-coupling monocular VINS’s loop detection on positioning accuracy improvement by calculating RMSE. Different datasets train the proposed SLAM system. The positioning accuracy is compared under the datasets with and without loop detection. Figure 12 compares the

<table>
<thead>
<tr>
<th>Number</th>
<th>Content</th>
<th>Model and parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Computer configuration</td>
<td>Intel i7-4510U</td>
</tr>
<tr>
<td>2</td>
<td>Memory size</td>
<td>8G</td>
</tr>
<tr>
<td>3</td>
<td>Operating system</td>
<td>ROS kinetic</td>
</tr>
<tr>
<td>4</td>
<td>CPU</td>
<td>2.66 GHz</td>
</tr>
<tr>
<td>5</td>
<td>Camera frame rate</td>
<td>20 HZ</td>
</tr>
<tr>
<td>6</td>
<td>Frequency of IMU</td>
<td>200 HZ</td>
</tr>
</tbody>
</table>

Figure 8: The initialization process of the SLAM system.

Figure 9: Structure of the proposed SLAM system based on tight-coupling monocular VINS.
influence of loop detection on the positioning accuracy of different datasets.

Figure 12 demonstrates that the RMSE of the dataset with loopback detection is smaller than that without loopback detection, and the gap between the two is large. The loop detection can play a certain role in reducing the error, and the effect is obvious. It can be used to solve the long-term error accumulation and ensure the stability of the long-term operation of the system. Therefore, the loop detection can effectively improve the positioning accuracy of the system.

3.3. Analysis of Attitude Estimation Test Results under Rotating Conditions. Under the condition of single rotation, the dynamic performance of the proposed SLAM system based on tight-coupling monocular VINS on the moving scene is tested. The dynamic experiment is conducted with the actual hardware and compared with the output results of the high-precision attitude solution module. Then, the pitch angle variation with time is analyzed. The estimated value is compared with the real value by considering the ATE. Figure 13 analyzes the variation of pitch angle and error curve.
Apparently, the estimated curve and the real curve of the pitch angle can be well fitted. The coincidence degree between the real value of the pitch angle and the estimated value under the proposed SLAM system by tight-coupling monocular VINS is high, and the AET can be controlled within $\pm 2$ with the change of time.

Furthermore, variation of roll angle with time is analyzed by comparing the estimated value with the real value and considering the AET. Figure 14 analyzes the variation of roll angle and error curve.

Figure 14 denotes that the fitting degree of the estimated curve and the real curve of the rolling angle are good. The coincident degree of the real value of the rolling angle and the estimated value by the proposed SLAM system by tight-coupling monocular VINS is high. The ATE can be controlled within $\pm 1.5^\circ$ with a change of time.

Afterward, the variation of yaw angle with time is analyzed by comparing the estimated value with the real value and the ATE. Figure 15 displays the variation of the yaw angle and error curve.

Figure 12: Influence of loop detection on the positioning accuracy of different datasets.

Figure 13: Variation of pitch angle and error curve: (a) the variation of pitch angle with time and (b) the pitch angle error curve.
Evidently, the fitting degree of the estimated curve and real curve of the yaw angle is good altogether. The true value of the yaw angle and estimated test fit well under the proposed SLAM system based on tight-coupling monocular VINS. The ATE can be controlled within ±2.5° with the change of time.

3.4. Error Analysis of Test Results of Different Algorithms. The designed algorithm is compared with the mainstream visual-inertial positioning algorithm Open Keyframe-based Visual-Inertial SLAM (OKVIS) and the RMSE obtained by a separate SLAM algorithm. Table 2 demonstrates the error analysis of the data set under different algorithms.

As exhibited in Table 2, the RMSE of each sequence obtained by the test results of this system is lower than that of OKVIS and SLAM. The error values of all data sequences are significantly lower than other algorithms. Therefore, it can be found that the accuracy of the designed algorithm is better than that of OKVIS and SLAM, and the RMSE is minimal.
To improve the positioning accuracy of the power inspection robot in different environments and the dynamic detection performance in moving scenes, on the basis of the IMU sensor measurement model, preintegration, and nonlinear optimization SLAM system theory, the shortcomings of the original SLAM system are improved, the optimization process of solving nonlinear SLAM based on GNM is established, and the SLAM algorithm structure of vision and IMU fusion under nonlinear optimization is constructed. Firstly, the performance is tested by moving scenes, and the results of the tight-coupling VINS algorithm are analyzed. Secondly, the influence of loop detection on positioning accuracy is explored, and finally, the test results of attitude estimation under rotation conditions are studied. By testing on the dataset, the results manifest that the roll angle error, pitch angle error, and yaw angle error in the robot attitude error are all within the controllable range. Position error, the three axes of X, Y, and Z are also within the error range, which can meet the positioning requirements of the power inspection robot. The results also show that loop detection can effectively improve the system’s positioning accuracy. In the attitude estimation test under pure rotation, the coincidence of the actual value and the predicted value of the robot attitude is good. With the change of time, the ATE can be controlled within ±2.5°. Compared with other algorithms, the designed algorithm has the smallest RMSE and has better accuracy for the positioning and navigation system of the power inspection robot. The disadvantage of the design method is that the local pose is estimated. In the future, it will face the challenge of continuous innovation of technical methods. It can be considered to introduce cutting-edge technologies and methods such as neural networks into the VINS. A further update is needed in the simulation scene setting to make it closer to the actual application scene. Moreover, sensors that provide measurements relative to the Earth coordinate system from the barometer can be fused to expand the application scenarios.

Data Availability

The data (data.xlsx) used to support the findings of this study have been deposited in the Baidu Netdisk repository. The link is https://pan.baidu.com/s/1HvAZEQ8a_ESmihd9107ASA. The password is ygg0.

4. Conclusion

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The study was supported by Shanghai Key Laboratory of Power Station Automation Technology (Grant no. 13DZZ2273800) Project.

References


