Article

Robust Tightly Coupled Pose Measurement Based on Multi-Sensor Fusion in Mobile Robot System

Gang Peng [1,2], Zezao Lu [1,2], Jiaxi Peng [1,2,*], Dingxin He [1,2], Xinde Li [1,3] and Bin Hu [4]

[1] Key Laboratory of Image Processing and Intelligent Control, Ministry of Education, Wuhan 430070, China; penggang@hust.edu.cn (G.P.); M201672539@hust.edu.cn (Z.L.); hedingxin@mail.hust.edu.cn (D.H.); xindeli@seu.edu.cn (X.L.)

[2] School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430070, China

[3] School of Automation, Southeast University, Nanjing 210096, China

[4] Shantui Construction Machinery Co., Ltd., Jining 272000, China; binhuyan@163.com

* Correspondence: m202072791@hust.edu.cn

Abstract: Currently, simultaneous localization and mapping (SLAM) is one of the main research topics in the robotics field. Visual-inertia SLAM, which consists of a camera and an inertial measurement unit (IMU), can significantly improve robustness and enable scale weak-visibility, whereas monocular visual SLAM is scale-invisible. For ground mobile robots, the introduction of a wheel speed sensor can solve the scale weak-visibility problem and improve robustness under abnormal conditions. In this paper, a multi-sensor fusion SLAM algorithm using monocular vision, inertia, and wheel speed measurements is proposed. The sensor measurements are combined in a tightly coupled manner, and a nonlinear optimization method is used to maximize the posterior probability to solve the optimal state estimation. Loop detection and back-end optimization are added to help reduce or even eliminate the cumulative error of the estimated poses, thus ensuring global consistency of the trajectory and map. The outstanding contribution of this paper is that the wheel odometer pre-integration algorithm, which combines the chassis speed and IMU angular speed, can avoid the repeated integration caused by linearization point changes during iterative optimization; state initialization based on the wheel odometer and IMU enables a quick and reliable calculation of the initial state values required by the state estimator in both stationary and moving states. Comparative experiments were conducted in room-scale scenes, building scale scenes, and visual loss scenarios. The results showed that the proposed algorithm is highly accurate—2.2 m of cumulative error after moving 812 m (0.28%, loopback optimization disabled)—robust, and has an effective localization capability even in the event of sensor loss, including visual loss. The accuracy and robustness of the proposed method are superior to those of monocular visual inertia SLAM and traditional wheel odometers.

Keywords: multi-sensor fusion; robot pose measurement; simultaneous localization and mapping; visual inertia system

1. Introduction

At present, the influential simultaneous localization and mapping (SLAM) systems with monocular vision include ORB-SLAM2 [1], LSD-SLAM [2], DSO [3], and SVO [4]. Among them, ORB-SLAM2 [4] has high adaptability and strong robustness to indoor and outdoor environments. However, the algorithm itself is based on pure vision. When there are no visual features, the accuracy and robustness of the pose estimation will decrease rapidly, and the algorithm may fail. Therefore, in the follow-up development of visual SLAM, in order to overcome the shortcomings of pure vision, the strategy of multisensor fusion is adopted.

Accordingly, sensors with scale measurement capabilities and monocular vision sensors are used to perform fusion vision SLAM to increase accuracy and robustness. Relatively stable and reliable solutions can be obtained with lasers [5]; however, at present, this method is generally only suitable for large-scale scenarios, such as unmanned driving, and is unsuitable for applications with limited costs. The inertial measurement unit (IMU) has become a generally accepted option. However, it exhibits a non-negligible cumulative error if run for a long period of time [6], especially in visually restricted conditions without texture or under weak illumination, in which case the visual mile cannot be used to correct the IMU error. In [7], the scale observability of a monocular visual inertial odometer on a ground mobile robot was analyzed in detail. When the robot moved at a constant speed, due to the lack of acceleration excitation, the constraint on the scale was lost, resulting in a gradual increase in the scale uncertainty and positioning error.

A wheel speed inertial odometer was integrated with a monocular visual odometer using the extended Kalman filter (EKF) [8,9], assuming the robot to be running on an ideal plane, and 3-DOF pose measurement was performed. The wheel speed inertial odometer uses wheel speed measurement for EKF status prediction; the angular speed measurement is integrated for dead reckoning; and the visual odometer method is used for an EKF measurement update. In the above-mentioned EKF-based loose coupling method, when the visual odometer cannot accurately calculate the pose due to insufficient visual characteristics, some proportion of its output in the filter will fall, resulting in ineffective visual observation. Consequently, the accuracy is reduced. VINS-Mono used tightly coupled nonlinear optimization [10,11]. By combining the IMU pre-integral measurement and visual measurement in a tightly coupled form to achieve the maximum posterior probability, we use the nonlinear optimization method to reach the optimal state. An open-source visual inertia SLAM algorithm VINS-Fusion [12,13] was developed based on VINS-Mono, supporting multiple sensor combinations (binocular camera + IMU; monocular camera + IMU; binocular camera only); GPS is used to improve the accuracy of the global path. In pursuit of better performance, wheel speed is used in SLAM system [14]—the wheel speed sensor and the visual odometer were integrated tightly. In this way, they make least-squares rather than optimization. However, they do not consider the unreliability of wheel speed. When a robot moves on uneven surfaces, or a wheel slip occurs, an incorrect wheel speed measurement will significantly affect the scale accuracy, and can potentially lead to system failure. In [15], a batch optimization framework with formulated edges was proposed. Moreover, tightly coupled nonlinear optimization methods were used to integrate vision, IMU, and wheel speed, and to perform pose. The error cost function was composed of vision errors, inertial measurement errors, and wheel speed sensor dead reckoning errors. In addition, assuming that the vehicle was moving on an approximate plane, a "soft" plane constraint term was added to the error cost function. While the robot is running with fixed acceleration or going straight, we do not know how long it moves in real time or what the direction is. Therefore, the encoder and soft plane constraints will make the task easier.

According to whether the visual feature points are added to the state vector of the filter or the optimization equation, the visual inertial odometry method based on filtering can be divided into two types: loose coupling and tight coupling. Further, the loose coupling method is divided into two parts: the visual odometer and state estimator. The visual odometer tracks the feature points and calculates the current camera pose, and the state estimator uses the IMU to measure and adjust the camera pose obtained by the visual odometer. As the visual features are invisible to the state estimator, they cannot be adjusted based on the IMU measurement, and the poor position accuracy of the visual features causes the accuracy of the visual odometer to decline. As a result, the correlation between measurements is not fully utilized for pose measurement, thereby reducing the accuracy. In the tight coupling method, the carrier pose and position of the feature point are used as variables to be estimated, and the filter method or optimized method is used for estimation.

Thus far, for the existing visual inertia SLAM algorithm, when the robot is moving at a constant speed, or purely rotating, and encounters scenes with insufficient visual features, the problems of low accuracy and poor robustness arise. Tightly coupled wheel speed information can help solve this problem. However, there are few studies on multi-sensor fusion SLAM for wheeled mobile robots based on vision, inertia, and wheel speed measurements that are tightly coupled and optimized. There is no complete solution similar to ORB-SLAM. Therefore, research on tightly coupled monocular visual odometer combined with wheel speed measurement is significant. Therefore, we use IMU and wheel odometry to make a monocular visual odometry system to gain the scale reconstruction.

The main novelty and contributions of this paper include:

1. A multi-sensor fusion SLAM algorithm using monocular vision, inertial measurement, and wheel speed measurement is proposed. The wheel speed and IMU angular velocity pre-integration of the wheel odometer can avoid the repeated integration caused by the linearization point change in the iterative optimization process.
2. Based on the state initialization of the wheel odometer and IMU, the initial state value required by the state estimator can be quickly and reliably calculated in both stationary and moving states.

The method in this paper solves the problem of poor pose estimation accuracy caused by the weak observability of monocular visual inertial SLAM, and further improves the robustness of pose estimation.

2. Multi-Sensor State Estimation Based on Tight Coupling Optimization

2.1. SLAM Based on Multi-Sensor Fusion

The multi-sensor fusion state estimator in this study uses monocular vision, IMU, and wheel odometer measurements based on feature point optical flow tracking [16]. None of these sensors can measure the absolute pose. Therefore, the multi-sensor fusion state estimator, as an odometer algorithm, has an unavoidable cumulative error. Therefore, we used key frame selection, loop detection [17], and back-end optimization algorithms based on the VINS-Mono framework and applied them to the multi-sensor fusion state estimator to form a complete SLAM system. Figure 1 shows a system block diagram of the SLAM method.

In a multi-sensor state estimation process, the main data processing and analysis processes include raw sensor input, data pre-processing (including calibration compensation, time alignment, and pre-fused wheel odometer), state initialization, state estimation problem solving, loopback detection, and backend optimization. The processes of optimal state estimator, wheel odometer pre-integration, and initialization reflect the contributions and innovations of this paper. The rest is our flexible use of existing methods. Sections 2.2–2.7 analyze the core multi-sensor fusion state estimation problem and the pre-integration process of IMU and wheel speed. Section 2.8 analyzes the initialization process of the state.
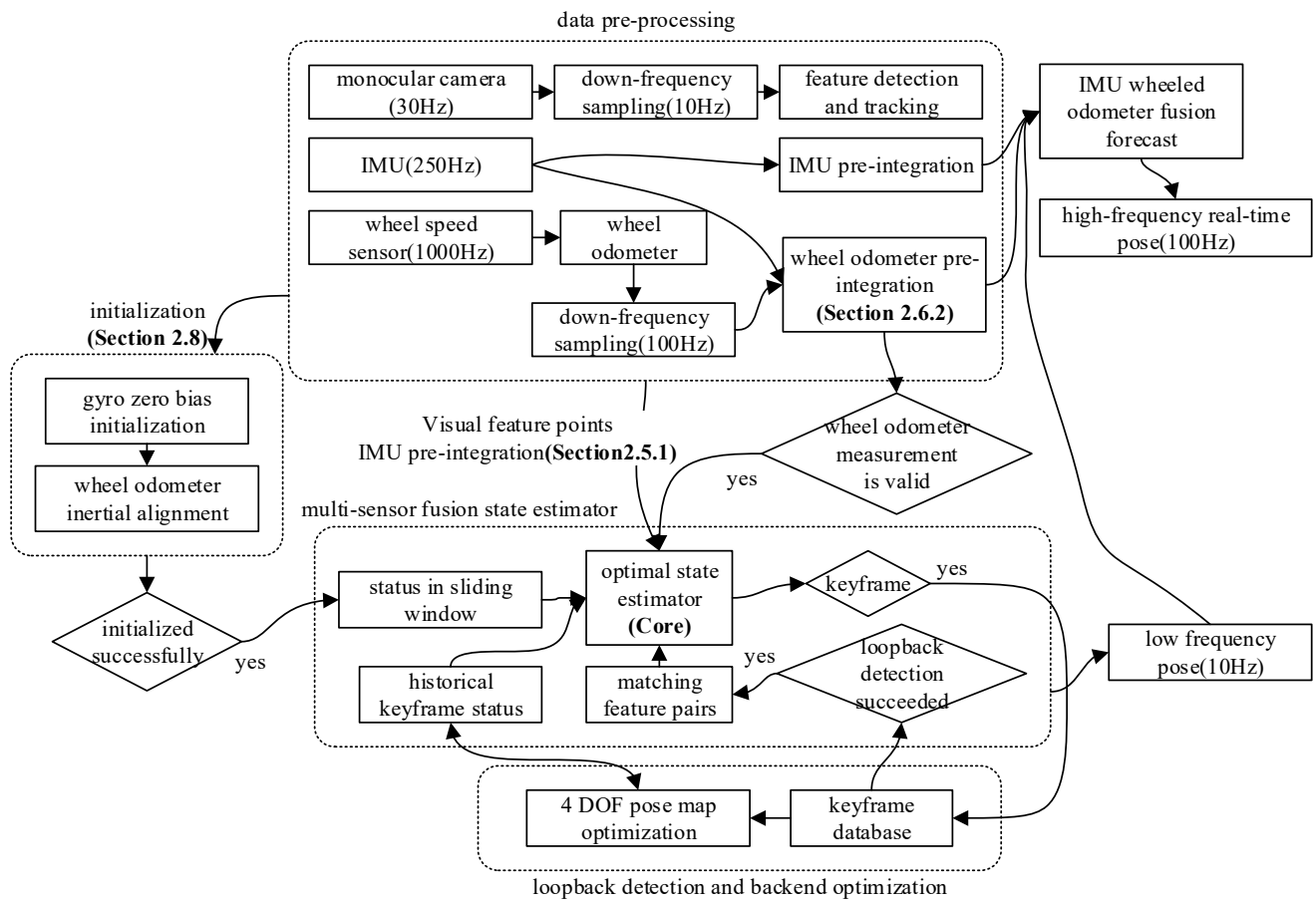
Figure 1. Block diagram of monocular vision inertial SLAM.

2.2. Variables to Be Estimated and Observation

Considering that the IMU zero offset always exists, when defining the variable $\mathcal{X}$ to be estimated, the IMU zero offset of each key frame needs to be included. Therefore, in the k key frame, we have:

$$\mathcal{X} = \{x_k, \lambda_l\}_{k \in \mathcal{K}, l \in \mathcal{L}}$$
$$x_k = \left[p_{B_k}^W, v_{B_k}^W, q_{B_k}^W, b_{ak}, b_{gk} \right] \tag{1}$$

For the IMU, $x_k \in R^{16 \times 1}$ is the state, $p_B^W \in R^{3 \times 1}$ denotes the position, $q_B^W \in R^{4 \times 1}$ is the attitude (quaternion form), $v_B^W \in R^{3 \times 1}$ denotes the speed, and $b_a \in R^{3 \times 1}$ and $b_g \in R^{3 \times 1}$ are the accelerometer and gyroscope bias, respectively. $\mathcal{K}$ is the key frames and $\mathcal{L}$ is the feature points. Further, λ_l is the inverse depth of $\mathcal{L}$ (the inverse of the Z-axis coordinate).

Additionally, pre-fusion wheel odometer observations are added based on VINS-Mono. The original wheel odometer is defined as:

$$\hat{m}_{odomk} \triangleq \left[\begin{array}{ccc} p_{xk} & p_{yk} & \theta_k \end{array} \right]^T \tag{2}$$

where $\{p_{xk}\ p_{yk}\}$ is the coordinate of the wheel odometer relative to the starting point and θ_k is the angle of rotation relative to the starting direction. Therefore, the observation $\mathcal{Z}$ used to constrain the $\mathcal{X}$:

$$\mathcal{Z} = \left\{ \mathcal{Z}_{C_i}, \mathcal{B}_{ij}, \mathcal{O}_{ij} \right\}_{(i,j) \in \mathcal{K}} \tag{3}$$

where visual feature point observation $\mathcal{Z}_{C_i} = \{\hat{z}_{il}\}_{l \in \mathcal{L}_i}$, containing all feature points $\mathcal{L}_i$ observed under the i key frame; $\mathcal{B}_{ij} = \{\hat{a}_t, \hat{\omega}_t\}_{t_i \leq t \leq t_j}$ is IMU pre-integration of accel-

erated speed and angular speed; and $\mathcal{O}_{ij} = \{\Delta\hat{m}_{odomt}, \hat{a}_{avgt}, \hat{\omega}_{avgt}\}_{t_i \leq t \leq t_j}$ is pre-fusion wheel odometer.

2.3. Optimal Estimation Problem

The maximum posterior estimation and Bayes' theorem shows:

$$\mathcal{X}^* = \underset{\mathcal{X}}{\operatorname{argmax}} p(\mathcal{Z}|\mathcal{X})p(\mathcal{X}) \tag{4}$$

Here, $p(\mathcal{Z}|\mathcal{X})$ is the conditional probability of observing the occurrence of $\mathcal{Z}$ under the existed state $\mathcal{X}$, which can be obtained by the observation equation and the covariance. $p(\mathcal{X})$ is the prior probability (edge probability) of $\mathcal{X}$. According to formula (3), we have:

$$
\begin{aligned}
p(\mathcal{Z}|\mathcal{X})p(\mathcal{X}) &= p(\mathcal{X}) \prod_{(i,j)\in\mathcal{K}} p(\mathcal{Z}_{C_i}, \mathcal{B}_{ij}, \mathcal{O}_{ij} \mid \mathcal{X}) \\
&= p(\mathcal{X}) \prod_{i\in\mathcal{K}} \prod_{l\in\mathcal{L}} p(\hat{z}_{il}|x_i, \lambda_l) \prod_{(i,j)\in\mathcal{K}} p(\{\hat{a}_t, \hat{\omega}_t\}_{t_i\leq t\leq t_j}|x_i, x_j) \\
&\times \prod_{(i,j)\in\mathcal{K}} p(\{\Delta\hat{m}_{odomt}, \hat{a}_{avgt}, \hat{\omega}_{avgt}\}_{t_i\leq t\leq t_j}|x_i, x_j)
\end{aligned}
\tag{5}
$$

We use the least-squares method to find the maximum posterior estimate, and use Mahalanobis distance to measure the difference between the residual and the covariance matrix.

$$
\begin{aligned}
\mathcal{X}^* &= \underset{\mathcal{X}}{\operatorname{argmin}} \Big\{ \|r_p - H_p\mathcal{X}\|^2 + \sum_{i\in\mathcal{K}}\sum_{l\in\mathcal{L}} \rho(\|r_C(\hat{z}_{il}, x_i, \lambda_l)\|^2_{\Sigma_{C_{il}}}) \\
&+ \sum_{(i,j)\in\mathcal{K}} \|r_{\mathcal{B}}(\{\hat{a}_t, \hat{\omega}_t\}_{t_i\leq t\leq t_j}, x_i, x_j)\|^2_{\Sigma_{\mathcal{B}_{ij}}} \\
&+ \sum_{(i,j)\in\mathcal{K}} \rho(\|r_{\mathcal{O}}(\{\Delta\hat{m}_{odomt}, \hat{a}_{avgt}, \hat{\omega}_{avgt}\}_{t_i\leq t\leq t_j}|x_i, x_j)\|^2_{\Sigma_{\mathcal{O}_{ij}}}) \Big\}
\end{aligned}
\tag{6}
$$

Here, $r\|(\cdot)\|^2_{\Sigma}$ is the Mahalanobis distance and Σ is the covariance matrix. $\{r_p, H_p\}$ is the prior information from marginalization; $r_C \in R$ is the visual residual; $r_{\mathcal{O}} \in R^{3\times1}$ is the wheel odometer residual; and $r_{\mathcal{B}} \in R^{15\times1}$ is the IMU residual. In addition, the Huber loss function ρ [18] is used to improve the robustness, where ρ is

$$\rho(s) = \left\{ \begin{array}{l} 1 \; s \geq 1 \\ 2\sqrt{s} - 1 \; s < 1 \end{array} \right. \tag{7}$$

2.4. Visual Measurement Constraints

The specific process is as follows: while the feature point l is first observed in the key frame i, it is recorded and tracked. Its spatial pose is defined as a function of the key frame i pose (p_{C_i}, q_{C_i}) and the inverse depth λ_l of the feature point. While the feature point l is observed again in the key frame j, a visual residual term is generated. The residual term $r_{C_{jl}}$ represents the error of the feature point l in the position of j. It is also called the re-projection error, which is a function of the key frame i pose (p_{C_i}, q_{C_i}).

$$r_C(\hat{z}_{jl}, \mathcal{X}) = r_{C_{jl}}(p^W_{B_i}, p^W_{B_j}, q^W_{B_i}, q^W_{B_j}, \lambda_l) = \begin{bmatrix} b_1 & b_2 \end{bmatrix}^T \cdot (\hat{\mathcal{P}}^{C_j}_l - \frac{\mathcal{P}^{C_j}_l}{\|\mathcal{P}^{C_j}_l\|}) \tag{8}$$

Here,

$$\hat{\mathcal{P}}^{C_j}_l = \pi_c^{-1}(\hat{z}_{il}) = \pi_c^{-1}\left(\begin{bmatrix} \hat{u}^{c_j}_l \\ \hat{v}^{c_j}_l \end{bmatrix}\right)$$

$$\mathcal{P}^{C_j}_l = (R^C_B(R^{B_j}_W(R^W_{B_i}(R^B_C \cdot \frac{1}{\lambda_l}\cdot\pi_c^{-1}\left(\begin{bmatrix} u^{c_u}_l \\ v^{c_u}_l \end{bmatrix}\right) + p^B_C) + p^W_{B_i} - p^W_{B_j}) - p^B_C) \tag{9}$$

In the formula, $R^{\mathcal{F}_1}_{\mathcal{F}_2} \in SO(3)$ represents the rotation matrix from $\mathcal{F}_2$ coordinate system to $\mathcal{F}_1$ coordinate system; $\hat{\mathcal{P}}^{c_j}_l \in R^{2\times 1}$ is the position where the feature point l is projected onto the unit ball in the key frame j; π_c^{-1} is the back projection function, which can project the pixel coordinates into the camera coordinate system C_j; $\mathcal{P}_l{}^{C_j} \in R^{2\times 1}$ is the position projected on the unit ball in the key frame j; to compare the error with $\mathcal{P}_l{}^{C_j}$, it needs to be transformed into the camera coordinate system C_j of the key frame j; and $\begin{bmatrix} b_1 & b_2 \end{bmatrix}$ are the two orthogonal base vectors on the tangent plane of the unit ball and the projection lines with the feature points in the orthogonal direction.

2.5. IMU Constraints

In the visual odometer method based on bundle adjustment, the state of the carrier is optimized and visual measurement is used to constrain. The IMU measurement between frames is added as a constraint on the optimization framework, thereby improving robustness.

2.5.1. IMU Pre-Integration

To reduce the complicated operation caused by reintegration, we used the IMU pre-integration method [19] to fuse IMU measurements between two consecutive key frames.

2.5.2. Residual Term

The IMU pre-integration processes the IMU measurement for a continuous period based on the given IMU zero offset and obtains the relative pose constraint between the initial and end states of the time period. The IMU pre-integration residual term is defined as:

$$
\begin{aligned}
&r_{\mathcal{B}}\left(\{\hat{a}_t, \hat{\omega}_t\}_{t_k \le t \le t_{k+1}}, x_k, x_{k+1}\right) \\
&= \begin{bmatrix} \delta\alpha^{B_k}_{B_k+1} \\ \delta\beta^{B_k}_{B_k+1} \\ \delta\theta^{B_k}_{B_k+1} \\ \delta b_{a}{}^{B_k}_{B_k+1} \\ \delta b_{g}{}^{B_k}_{B_k+1} \end{bmatrix} = \begin{bmatrix} R\{q^W_{B_k}\}^T(p^W_{B_{k+1}} - p^W_{B_k} - v^W_{B_k}\Delta t + \frac{1}{2}g\Delta t^2) - \hat{\alpha}^{B_k}_{B_k+1} \\ R\{q^W_{B_k}\}^T(v^W_{B_{k+1}} - v^W_{B_k} + g\Delta t) - \hat{\beta}^{B_k}_{B_k+1} \\ 2\{\hat{q}^{B_k}_{B_k+1} * \otimes q^W_{B_k} * \otimes q^W_{B_{k+1}}\}_{xyz} \\ b_{aB_{k+1}} - b_{aB_k} \\ b_{gB_{k+1}} - b_{gB_k} \end{bmatrix}
\end{aligned} \tag{10}
$$

The random distribution of the residual term $r_{\mathcal{B}}$ conforms to $N(0, \Sigma_B)$ and Σ_B is obtained by updating the covariance equation. $\delta\theta^{B_k}_{B_k+1} \in R^{3\times 1}$ is three-dimensional small perturbation and δ represents the error term. The IMU pre-integration provides constraints on the variables to be optimized contained in the two key frames before and after. In the process of nonlinear optimization, the essence of the constraint is to provide the direction and gradient of the variable to be optimized by calculating the Jacobian matrix of the residual. As the direction of gravity is obtained during the initialization of the visual inertial odometer, the gravity acceleration g^W is not used as a variable to be optimized.

2.6. Wheeled Odometer Constraints

On ground mobile robots, a wheel speed meter is typically used to perform dead reckoning to obtain continuous relative poses of the robot. The continuous position and reliable scale estimation of the wheel odometer make it suitable for tasks such as path planning and navigation.

2.6.1. Two-Dimensional Wheel Odometer Algorithm

The two-dimensional wheel odometer has an unavoidable cumulative, error but can provide a continuous carrier trajectory. As the wheel speed meter measures the average wheel speed over a period of time, the chassis speed measurement $\hat{m}_{base}$ measures the average speed during this time. The position and attitude update methods of the wheel speed odometer include Euler points, median points, and higher-order Runge–Kutta methods. As

the sampling speed of the wheel speed meter is high (1 kHz), the Euler integration method is used to reduce the calculation time of the main control microcontroller. This is achieved while assuming that the chassis moves with fixed direction and speed in the original direction during the period and rotates to a new direction at the end of the time period.

The initial state of the wheel odometer is $\hat{m}_{odom0} = \begin{bmatrix} p_{x0} & p_{y0} & \theta_0 \end{bmatrix}^T = \mathbf{0}$. Given the previous state $\hat{m}_{odomk-1} = \begin{bmatrix} p_{xk-1} & p_{yk-1} & \theta_{k-1} \end{bmatrix}^T$ of the wheel odometer, the current chassis speed measurement $\hat{v}_{basek} = \begin{bmatrix} v_{xk} & v_{yk} & \omega_k \end{bmatrix}^T$, and the time difference $dt_k = t_k - t_{k-1}$, we can obtain the new wheel odometer state $\hat{m}_{odomk}$ as:

$$\hat{m}_{odomk} = \begin{bmatrix} p_{xk} & p_{yk} & \theta_k \end{bmatrix}^T \tag{11}$$

$$\begin{bmatrix} p_{xk} \\ p_{yk} \end{bmatrix} = \begin{bmatrix} p_{xk-1} \\ p_{yk-1} \end{bmatrix} + \begin{bmatrix} cos\theta_{k-1} & -sin\theta_{k-1} \\ sin\theta_{k-1} & cos\theta_{k-1} \end{bmatrix} \begin{bmatrix} v_{xk} \\ v_{yk} \end{bmatrix} dt_k \tag{12}$$

$$\theta_k = \theta_{k-1} + \omega_k dt_k \tag{13}$$

2.6.2. Wheel Odometer Pre-Integration

In this paper, we propose a pre-integration method for wheel odometer. We use IMU and wheel speed to measure the relative pose between two key frames. The data of the two sensors makes pre-fusion $\hat{m}_{fused\ odom}$. Thereafter, only the $\hat{m}_{fused\ odom}$ is measured, according to the wheel odometer kinematics equation, and a continuous calculation and integration is made to find the displacement.

The incremental update equation of the wheel odometer is:

$$\begin{cases} p_{O_{i+1}}^{O_k} = p_{O_i}^{O_k} + R\{q_{O_i}^{O_k}\}\Delta p_{O_{i+1}}^{O_i} = p_{O_i}^{O_k} + R\{q_{O_i}^{O_k}\}(\Delta \hat{p}_{O_{i+1}}^{O_i} + D_{O_{k+1}}^{O_k}\eta_{os}) \\ q_{O_{i+1}}^{O_k} = q_{O_i}^{O_k} \otimes q_{O_{i+1}}^{O_i} = q_{O_i}^{O_k} \otimes q\{R_B^O(\hat{\omega}_{avgi+1} - b_g - \eta_g)\Delta t\} \\ b_{gi+1} = b_{gi} + \eta_{b_g}\Delta t \end{cases} \tag{14}$$

where $\Delta \hat{p}_{O_{i+1}}^{O_i} \triangleq \begin{bmatrix} \Delta \hat{p}_{xi+1} & \Delta \hat{p}_{yi+1} & 0 \end{bmatrix}^T$ denotes wheel odometer with noise measurement; the initial state value $p_0 = 0$, $q_0 = [1\ 0\ 0\ 0]^T$. $x = \begin{bmatrix} p^{O_k} & q^{O_k} & b_g \end{bmatrix}^T$ is the pre-credit term.

The nominal weight of the wheel odometer pre-integration item can be incrementally updated based on the pre-fusion wheel odometer measurement:

$$\begin{cases} \hat{p}_{O_{i+1}}^{O_k} = \hat{p}_{O_i}^{O_k} + R\{\hat{q}_{O_i}^{O_k}\}\Delta \hat{p}_{O_{i+1}}^{O_i} \\ \hat{q}_{O_{i+1}}^{O_k} = \hat{q}_{O_i}^{O_k} \otimes \hat{q}_{O_{i+1}}^{O_i} = \hat{q}_{O_i}^{O_k} \otimes \hat{q}\{R_B^O(\hat{\omega}_{avgi+1} - b_g)\Delta t\} \\ \hat{b}_{gi+1} = \hat{b}_{gi} \end{cases} \tag{15}$$

The initial value of nominal weight: $\hat{p}_0 = 0$, $\hat{q}_0 = [1\ 0\ 0\ 0]^T$.

According to the definition of the error amount of the pre-integration term, an incremental update formula of the error amount of the pre-integration term is:

$$\begin{cases} \delta p_{O_{i+1}}^{O_k} = \delta p_{O_i}^{O_k} + R\{q_{O_i}^{O_k}\}D_{O_{k+1}}^{O_k}\eta_{os} \\ \delta \theta_{O_{i+1}}^{O_k} = \delta \theta_{O_i}^{O_k} - (\hat{\omega}_{avgi+1} - b_g)\hat{\delta \theta}_{O_i}^{O_k}\Delta t - \delta b_g \Delta t + \eta_g \Delta t\} \\ \delta b_{gi+1} = \delta b_{gi} + \eta_{b_g}\Delta t \end{cases} \tag{16}$$

$\hat{(\cdot)}$ is 3×3 anti-symmetric matrix of Lie algebra SO (3). The nominal value of the wheel odometer pre-integration term depends on the pre-fusion wheel odometer measurement and the gyroscope zero offset. For the variable to be optimized, the zero bias of the gyroscope needs to be continuously adjusted in the pose measurement process to reduce the residual error. Therefore, in the optimization process, the partial derivatives $J_{b_g}^p$ and $J_{b_g}^\theta$

of the nominal value of the pre-integral term of the wheel odometer with respect to the zero offset of the gyroscope need to be used.

According to the incremental update of the error value of the pre-integration term of the wheel odometer, the Jacobian matrix of the error value between the two frames before and after can be obtained as:

$$J_{\delta x_i}^{\delta x_{i+1}} = \frac{\delta x_{i+1}}{\delta x_i} = I + F_i \tag{17}$$

According to the definition of the nominal value of the pre-integration item of the wheel odometer, the Jacobian matrix of the error value is the nominal Jacobian matrix of $J_{\hat{x}_i}^{\hat{x}_{i+1}} = J_{\delta x_i}^{\delta x_{i+1}}$. Therefore, according to the chain-derivation rule, $J_{\hat{x}_0}^{\hat{x}_{i+1}} = J_{\hat{x}_i}^{\hat{x}_{i+1}} J_{\hat{x}_0}^{\hat{x}_i}$, the update equation of the nominal value of the pre-integration term of the wheel odometer with respect to the zero-biased Jacobian matrix is:

$$J_{\hat{x}_0}^{\hat{x}_{i+1}} = (I + F_i) J_{\hat{x}_0}^{\hat{x}_i} \tag{18}$$

The initial value of the Jacobian matrix: $J_0 = J_{\hat{x}_0}^{\hat{x}_0} = diag(9)$.

2.6.3. Residual Term

Definition: In the least-squares problem of robot pose measurement, the wheel odometer residual term $r_{\mathcal{O}}$ represents the error distance between the frame-to-frame relative displacement $\hat{p}_{O_{k+1}}^{O_k}$ and the key frame displacement $p_{O_{k+1}}^{O_k}$ in the variable to be optimized, where $\hat{p}_{O_{k+1}}^{O_k}$ is the observation, and $p_{O_{k+1}}^{O_k}$ is the estimator.

$$r_{\mathcal{O}}(\{\Delta \hat{m}_{odomt}, \hat{\omega}_{avgt}\}_{t_k \le t \le t_{k+1}}, x_k, x_{k+1}) = \left[\delta p_{O_{k+1}}^{O_k} \right] = p_{O_{k+1}}^{O_k} - \hat{p}_{O_{k+1}}^{O_k} \tag{19}$$

The wheel odometer residual does not include the errors $\delta \theta_{O_{k+1}}^{O_k}$ and $\delta b_{g O_{k+1}}^{O_k}$ with respect to the rotation and gyro zero offset. This is as these terms are already defined in the residual term of the IMU pre-integration. The IMU pre-integration uses the original IMU measurement as the angular velocity input, which provides higher rotational integration accuracy than the wheel odometer pre-integration measured with a lower frequency pre-fused wheel odometer.

To use the variable x_k, x_{k+1} to represent the wheel odometer pre-integration residual term, $p_{O_{k+1}}^{O_k}$ needs to be transformed:

$$
\begin{aligned}
p_{O_{k+1}}^{O_k} &= R_W^{O_k}(p_{O_{k+1}}^W - p_{O_k}^W) \\
&= R_O^{B-1} R\{q_{B_k}^{W-1}\}(p_{B_{k+1}}^W - p_{B_k}^W) + R_O^{B-1} R\{q_{B_k}^{W-1}\} R\{q_{B_{k+1}}^W\} p_O^B - R_O^{B-1} p_O^B
\end{aligned} \tag{20}
$$

We obtain the residual term expressed using only the variables to be optimized and the wheel odometer pre-integration:

$$r_{\mathcal{O}} = R_O^{B-1} R\{q_{B_k}^{W-1}\}(p_{B_{k+1}}^W - p_{B_k}^W) + R_O^{B-1} R\{q_{B_k}^{W-1}\} R\{q_{B_{k+1}}^W\} p_O^B - R_O^{B-1} p_O^B - \hat{p}_{O_{k+1}}^{O_k} \tag{21}$$

Here, R_O^B and p_O^B are known constants.

As a maximum posterior problem, the robot pose measurement is the same as least-squares by introducing a covariance matrix of the residuals to transform the residuals with dimensions into a unified probability representation. The wheeled odometer residual $r_{\mathcal{O}}$ obeys the covariance matrix $\Sigma_{\mathcal{O}}$ of the wheeled odometer pre-integration, $r_{\mathcal{O}} \sim N(0, [\Sigma_{\mathcal{O}}]_p)$. Here, $[\Sigma_{\mathcal{O}}]_p$ represents the displacement covariance in the wheel odometer pre-integration covariance matrix $\Sigma_{\mathcal{O}}$, $[\Sigma_{\mathcal{O}}]_p = [\Sigma_{\mathcal{O}}]_{left\ right\ 3 \times 3}$.

Regarding the Jacobian matrix, according to the definition of the wheel odometer residual $r_{\mathcal{O}}$, in the optimization process, the residual value $r_{\mathcal{O}}$ will change with the adjustment of the previous key frame poses $p_{B_k}^W$ and $q_{B_k}^W$, and the poses $p_{B_{k+1}}^W$ and $q_{B_{k+1}}^W$ of the next

key frame, and the gyroscope zero offset b_{gi} of the previous frame. To provide the necessary gradient direction for optimization, the system needs to be linearized in the current state x_k, x_{k+1} and the ratio between the increment of the residual $\partial r_\mathcal{O}$ and the increment of the variable to be optimized is calculated. Thus, the Jacobian matrix $J^{r_\mathcal{O}}_{x_k, x_{k+1}}$ is defined:

$$\partial r_\mathcal{O} = J^{r_\mathcal{O}}_{x_k, x_{k+1}} \begin{bmatrix} \partial x_k \\ \partial x_{k+1} \end{bmatrix} \tag{22}$$

$$\begin{bmatrix} \delta p^{O_k}_{O_{k+1}} \end{bmatrix} = J^{r_\mathcal{O}}_{x_k, x_{k+1}} \begin{bmatrix} \delta p^W_{B_k} \\ \delta v^W_{B_k} \\ \delta \theta^W_{B_k} \\ \delta b_{a B_k} \\ \delta b_{g B_k} \\ \delta p^W_{B_k} \\ \delta v^W_{B_k} \\ \delta \theta^W_{B_k} \\ \delta b_{a B_k} \\ \delta b_{g B_k} \end{bmatrix} \tag{23}$$

Here, $J^{r_\mathcal{O}}_{x_k, x_{k+1}} \in R^{3 \times 30}$. As the increment is small, using the quaternion definition will produce additional degrees of freedom. The increment for the rotation state in the formula is defined as the shaft angle representation.

As the wheel odometer residual is only related to some variables in the previous key frame state x_k and next key frame state x_{k+1}, the value of the Jacobian matrix $J^{r_\mathcal{O}}_{x_k, x_{k+1}}$ is:

$$J^{r_\mathcal{O}}_{x_k, x_{k+1}} = \begin{bmatrix} \dfrac{\partial \delta p^{O_k}_{O_{k+1}}}{\partial \delta p^W_{B_k}} & 0 & \dfrac{\partial \delta p^{O_k}_{O_{k+1}}}{\partial \delta \theta^W_{B_k}} & 0 & \dfrac{\partial \delta p^{O_k}_{O_{k+1}}}{\partial \delta b_{gi}} & \dfrac{\partial \delta p^{O_k}_{O_{k+1}}}{\partial \delta p^W_{B_{k+1}}} & 0 & \dfrac{\partial \delta p^{O_k}_{O_{k+1}}}{\partial \delta \theta^W_{B_{k+1}}} & 0 & 0 \end{bmatrix} \tag{24}$$

$$\begin{aligned} \dfrac{\partial \delta p^{O_k}_{O_{k+1}}}{\partial \delta p^W_{B_k}} &= -R^{B-1}_O R\{q^{W-1}_{B_k}\} \\ \dfrac{\partial \delta p^{O_k}_{O_{k+1}}}{\partial \delta \theta^W_{B_k}} &= -R^{B-1}_O \left[R\{q^{W-1}_{B_k}\}(p^W_{B_{k+1}} - p^W_{B_k} + R\{q^W_{B_{k+1}}\}p^B_O) \right]^{\wedge} \\ \dfrac{\partial \delta p^{O_k}_{O_{k+1}}}{\partial \delta b_{gi}} &= -J^p_{b_g} \\ \dfrac{\partial \delta p^{O_k}_{O_{k+1}}}{\partial \delta p^W_{B_{k+1}}} &= R^{B-1}_O R\{q^{W-1}_{B_k}\} \\ \dfrac{\partial \delta p^{O_k}_{O_{k+1}}}{\partial \delta \theta^W_{B_{k+1}}} &= -R^{B-1}_O R\{q^{W-1}_{B_k}\} R\{q^W_{B_{k+1}}\}(p^B_O)^{\wedge} \end{aligned} \tag{25}$$

2.7. Marginalization and Prior Constraints

The state of the key frame and its related observations are constantly removed from the optimization equation. If all observations related to the removed key frames are directly discarded, the constraints of state estimation will be reduced, and the loss of valid information will lead to a decrease in accuracy. Here, a marginalization algorithm is used, while removing the key frames, retaining the removed observations to constrain the optimization variables. According to [6], the use of the Gauss–Newton method can be understood as adding an increment to the variable to be optimized; the objective function is the smallest.

If the residual function $r(x)$ is linearized at x, and the Jacobian matrix J^r_x is obtained, the nonlinear least-squares problem becomes a linear least-squares problem:

$$\min_{\delta x} \|r(x + \delta x)\|^2 \implies \min_{\delta x} \|r(x) + J^r_x \delta x\|^2 \tag{26}$$

Here, $\|r(x) + J_x^r \delta x\|^2 = [r(x) + J_x^r \delta x]^T [r(x) + J_x^r \delta x]$. Taking the derivative of this formula with respect to δx be 0, we obtain:

$$J_x^{r\,T} J_x^r \delta x = -J_x^{r\,T} r(x) \tag{27}$$

Let $H = J_x^{r\,T} J_x^r$, $b = -J_x^{r\,T} r(x)$; thus, we obtain the incremental equation $H\delta x = b$, where H is called the Hessian matrix.

Divide the variable x to be optimized into the part x_a, which needs to be removed, and the part x_b, $\delta x = [\delta x_a \ \delta x_b]^T$, which needs to be retained, then the incremental equation becomes:

$$\begin{bmatrix} H_a & H_b \\ H_b^T & H_c \end{bmatrix} \begin{bmatrix} \delta x_a \\ \delta x_b \end{bmatrix} = \begin{bmatrix} b_a \\ b_b \end{bmatrix} \tag{28}$$

The Schur method is used to eliminate the element to obtain the solution of δx_b:

$$\begin{bmatrix} I & 0 \\ -H_b^T H_a^{-1} & I \end{bmatrix} \begin{bmatrix} H_a & H_b \\ H_b^T & H_c \end{bmatrix} \begin{bmatrix} \delta x_a \\ \delta x_b \end{bmatrix} = \begin{bmatrix} I & 0 \\ -H_b^T H_a^{-1} & I \end{bmatrix} \begin{bmatrix} b_a \\ b_b \end{bmatrix} \tag{29}$$

$$\begin{bmatrix} H_a & H_b \\ 0 & H_c - H_b^T H_a^{-1} H_b \end{bmatrix} \begin{bmatrix} \delta x_a \\ \delta x_b \end{bmatrix} = \begin{bmatrix} b_a \\ b_b - H_b^T H_a^{-1} b_a \end{bmatrix} \tag{30}$$

Intercepting the second row of the above matrix, we obtain:

$$(H_c - H_b^T H_a^{-1} H_b)\delta x_b = b_b - H_b^T H_a^{-1} b_a \tag{31}$$

In the above formula, only x_b is unknown, and no information in H and b is lost. This process removes the rows and columns related to x_a from the incremental equation, marginalizes the state x_a that needs to be removed, and retains the historical observation constraints on the state x_b. When the next image frame arrives, the prior information in the above Formula (31) will be used as a prior constraint term to construct a nonlinear least-squares problem.

2.8. State Initialization

VINS-Mono uses multiple steps to initialize the state: gyroscope zero offset correction, initializing gravity, speed, and scale coefficients, and modifying the direction of gravity. The disadvantage of this method is that it depends on sufficient visual measurement of parallax and sufficient acceleration excitation. When there is no abnormal situation, such as skidding, the wheel odometer has better accuracy and reliability over a short distance and a short duration. Compared with monocular vision, there is no scale uncertainty, and it is easier to initialize the keyframe pose, velocity, and gravity directions.

2.8.1. Gyro Zero Offset Initialization

As the gyroscope and wheel odometer measurements are on the same rigid body, the rotations of the two are the same. The relative rotation between the two key frames can be obtained through IMU pre-integration and wheel odometer pre-integration, respectively: $q_{B_{k+1}}^{B_k}$ and $q_{O_{k+1}}^{O_k}$. The rotation term of the pre-integration of the wheeled odometer above is also obtained through the gyro integration and has no reference value. Therefore, during the initialization process, the gyroscope pre-integration will use the heading angle for rotation integration. Rotation term $q_{B_{k+1}}^{B_k}$ is a function of the gyroscope bias $q_{B_{k+1}}^{B_k}$. If the error between $q_{B_{k+1}}^{B_k}$ and $q_{O_{k+1}}^{O_k}$ is used as a constraint, the gyroscope bias b_g can be estimated. Assuming that the gyro bias b_{gk} of each key frame during the initialization process is the same $b_{gk} = b_g$, the construction of the least-squares problem is as follows:

$$b_g^* = \underset{b_g}{\arg\min} \sum_{i=1}^{k-1} \|q_O^B \otimes \hat{q}_{O_{i+1}}^{O_i}{}^{-1} \otimes q_B^O \otimes q_{B_{k+1}}^{B_k}\|^2 \tag{32}$$

Linearizing the rotation transform at $\hat{q}_{B_{k+1}}^{B_k}$, we obtain:

$$q_{B_{k+1}}^{B_k} = \hat{q}_{B_{k+1}}^{B_k} \otimes q\{J_{b_g}^q b_g\} \approx \hat{q}_{B_{k+1}}^{B_k} \otimes \begin{bmatrix} 1 \\ \frac{1}{2} J_{b_g}^q b_g \end{bmatrix} \tag{33}$$

Here, $J_{b_g}^q$ is the partial derivative of the inter-frame rotation $q_{B_{k+1}}^{B_k}$. b_g is the gyroscope zero bias. The objective function of the least-squares problem is written as:

$$q_O^B \otimes \hat{q}_{O_{i+1}}^{O}{}^{-1} \otimes q_B^O \otimes q_{B_{k+1}}^{B_k} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\hat{q}_{B_{k+1}}^{B_k} \otimes \begin{bmatrix} 1 \\ \frac{1}{2} J_{b_g}^q b_g \end{bmatrix} = q_O^B \otimes \hat{q}_{O_{i+1}}^{O} \otimes q_B^O \tag{34}$$

$$\begin{bmatrix} 1 \\ \frac{1}{2} J_{b_g}^q b_g \end{bmatrix} = \hat{q}_{B_{k+1}}^{B_k}{}^{-1} \otimes q_O^B \otimes \hat{q}_{O_{i+1}}^{O} \otimes q_B^O$$

Considering only the imaginary part of the quaternion, we obtain:

$$J_{b_g}^q b_g = 2\{\hat{q}_{B_{k+1}}^{B_k}{}^{-1} \otimes q_O^B \otimes \hat{q}_{O_{i+1}}^{O} \otimes q_B^O\}_{xyz} \tag{35}$$

The above Formula (35) conforms to the format of Hx = b, and the Cholesky decomposition can be used to find the least-squares solution:

$$J_{b_g}^{q\ T} J_{b_g}^q b_g = 2 J_{b_g}^{q\ T} \{\hat{q}_{B_{k+1}}^{B_k}{}^{-1} \otimes q_O^B \otimes \hat{q}_{O_{i+1}}^{O} \otimes q_B^O\}_{xyz} \tag{36}$$

If the robot rotates rapidly during the gyro work offset initialization process, the elasticity of the wheel, the rigid connection between the wheel and the IMU, the misalignment of the wheel odometer clock and the IMU clock, and a calibration error of the wheel odometer rotation scale factor may lead to poor gyro work offset initialization results.

2.8.2. Initialization of Key Frame Speed and Gravity

As the Mecanum wheel will tremble during movement, and the wheeled odometer algorithm can only obtain the heading angle information, it is difficult to obtain accurate relative rotation between key frames through wheeled odometer integration. In the previous step, the zero offset of the gyroscope has been initialized, and the relative rotation between all key frames can be obtained through IMU pre-integration. As the rotation is known, the key frame speed and gravity can be calculated by solving linear equations. Decomposing the position term $\alpha_{B_{k+1}}^{B_k}$ and speed term $\beta_{B_{k+1}}^{B_k}$ in the IMU pre-integration and transforming it into the form of matrix multiplication $\hat{z}_{B_{k+1}}^{B_k} = H_{B_{k+1}}^{B_k} x_{B_{k+1}}^{B_k}$, we obtain:

$$\begin{bmatrix} \alpha_{B_{k+1}}^{B_k} - p_O^B + R_{B_{k+1}}^{B_k} p_O^B \\ \beta_{B_{k+1}}^{B_k} \end{bmatrix} = \begin{bmatrix} -I\Delta t & 0 & \frac{1}{2} R_{B_0}^{B_k} \Delta t^2 & R_O^B \hat{p}_{O_{k+1}}^{O_k} \\ -I & R_{B_{k+1}}^{B_k} & R_{B_0}^{B_k} \Delta t & 0 \end{bmatrix} \begin{bmatrix} v_{B_k}^{B_k} \\ v_{B_{k+1}}^{B_{k+1}} \\ g^{B_0} \\ s \end{bmatrix} \tag{37}$$

Here, $\hat{z}_{B_{k+1}}^{B_k}$ is the IMU pre-integration measurement between the key frames B_k and B_{k+1}. The variable to be estimated related to the key frames B_k and B_{k+1} is defined as $x_{B_{k+1}}^{B_k}$, and $H_{B_{k+1}}^{B_k}$ represents the constraint between the measurement $\hat{z}_{B_{k+1}}^{B_k}$ and the $x_{B_{k+1}}^{B_k}$. The s in the $x_{B_{k+1}}^{B_k}$ represents the distance of the wheel odometer with respect to the actual distance, that is, the X-axis and Y-axis scale factors of the wheel odometer S_x, S_y. If the IMU excitation is sufficient, it can be used to calibrate the scale factor. S = 1 is defined here for the reliability of initialization.

To reduce initialization errors and improve reliability, multiple key frame measurements need to be used as constraints to calculate the key frame speed and gravity direction. By combining the multiple linear equations above, we can obtain the least-squares problem:

$$x^* = \operatorname*{argmin}_{x} \sum_{k \in Frames} \| \hat{z}_{B_{k+1}}^{B_k} - H_{B_{k+1}}^{B_k} x_{B_{k+1}}^{B_k} \|^2, x = \begin{bmatrix} v_{B_0}^{B_0} \\ v_{B_1}^{B_1} \\ \vdots \\ v_{B_k}^{B_k} \\ g^{B_0} \\ s \end{bmatrix} \tag{38}$$

Here, x is the variable to be estimated, and x^* is the optimal estimated value of x. The program uses Cholesky decomposition to solve the least-squares problem:

$$\mathrm{H}^T H x = H^T \hat{z} \tag{39}$$

In the formula, the matrix H is obtained by inserting all $H_{B_{k+1}}^{B_k}$ into empty columns at the corresponding positions of the unrelated variables and summing them, and $\hat{z}$ is obtained by combining all $\hat{z}_{B_{k+1}}^{B_k}$.

3. Experiments

In this paper, a mobile robot platform is built, using a dual-processor architecture. The high-performance PC upper computer is used for data processing of each sensor and the SLAM algorithm operation, and the embedded controller lower computer is used for the motion control of the mobile robot chassis. The parameters of the high-performance PC are Intel Core i7-9750H CPU @ 2.6 GHz, 16 GB RAM. The camera is IntelRealSense ZR300.

3.1. Accuracy Verification Experiment

3.1.1. Room-Scale Pose Measurement Experiment

Experimental conditions: In a laboratory where objects are placed in a complex environment, as shown in Figure 2, the control robot walked through all the channels. The channel width is narrow, and the width at the narrowest point is less than 1 m; the movement speed was maintained at approximately 0.5 m/s; a large number of fixed marking points are arranged inside the room to obtain the real position of the robot in order to analyze the positioning error of the robot. Abnormal conditions during the experiment included:

3. magnetic guide bars with a height of approximately 0.5 cm were fixed on the ground, and the wheels slipped slightly as they passed;

4. due to turning too close to the weakly textured wall surface, the visual tracking was completely lost several times.
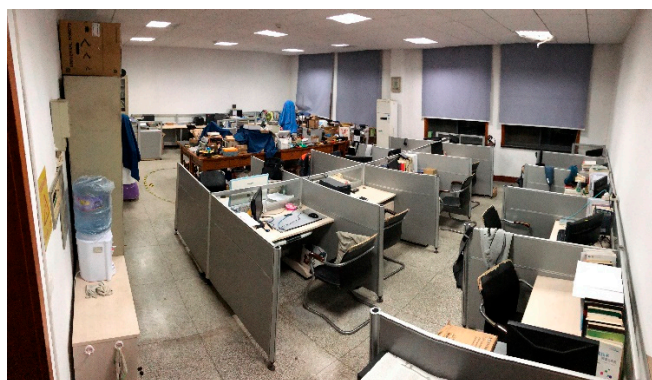

Figure 2. Room-scale experimental environment.

The RViz interface at the end of the test data playback is shown in Figure 3, which includes the raw picture, feature points, matched feature points, and the path of pose measurement.
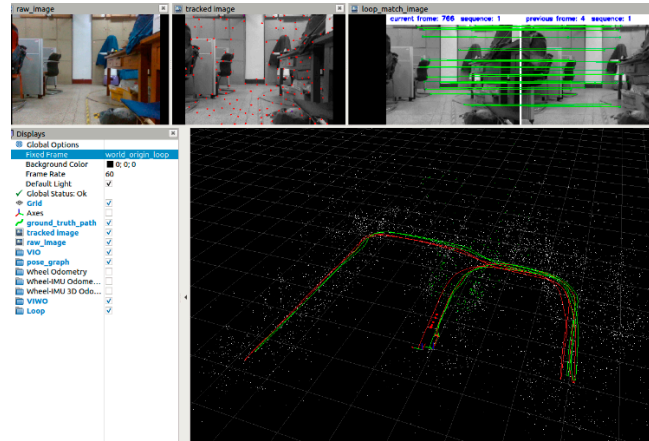

**Figure 3.** RViz interface at the end of test data playback.

The path diagrams of pose measurement, as shown in Figure 4, start along the positive X-axis. Therefore, the distance between the start point and end point shows the experimental accuracy. Throughout the experiment, the car traveled for 184.3 s, the average speed was 0.264 m/s, the maximum speed was 0.591 m/s, the cumulative translation was 51.321 m, the cumulative rotation was 3428.318°, and the chassis had no abnormalities. As this experiment was performed in an actual scene, the experimental system did not have the equipment to measure ground truth; therefore, it could only measure the distance error of the car passing several fixed road marking points. In each experiment, the robot was controlled to accurately pass the road marking point and keep the rotation direction consistent. Therefore, the pose error in Table 1 is average position error and yaw angle error among several road marking points. When the robot passes through each road marking point, a timestamp is marked. Calculate the coordinate error (ΔX and ΔY), position error ($\Delta P(m)$, absolute value), and yaw angle error (ΔA) between the algorithm trajectory pose and the road marking point when it is at the time stamp. For each algorithm, take the average of the errors generated by all road marking points as the final experimental result. The position error rate ($\Delta P(\%)$) is the ratio of the position error to the cumulative translation. The table shows our algorithm performed best, with and without closed loops. The position error was 0.206 m and 0.015 m, respectively. As shown in Figure 4, we can clearly see the error of the different methods.

Table 1. Room-scale pose measurement results.

Algorithm	ΔX (m)	ΔY (m)	ΔP (m)	ΔP (%)	ΔA (°)
wheel	−0.785	−0.389	0.87	1.71	1.907
wheel inertial	−0.905	−0.040	0.906	1.77	−0.507
VINS-Mono	0.851	−0.220	0.879	1.71	−0.574
VINS-Mono (loop)	0.009	0.020	0.022	0.04	−0.530
ours	0.148	−0.143	0.206	0.40	−0.213
ours (loop)	−0.003	0.015	0.015	0.03	−0.443

As listed in Table 1, we compared the error of the X-axis, Y-axis, position, and heading angle. According to the data presented in Table 1, the accuracy of posture estimation using the monocular vision inertial wheel odometer algorithm was better than that of the VINS-Mono algorithm, and the accumulated position error was only approximately 0.2 m after

51 m. We divided the cumulative position error by the cumulative translation distance to obtain the cumulative position error rate. The position error rate of the proposed algorithm was only 0.4%, which was lower than that of VINS-Mono (1%). This experiment verified that a multi-sensor fusion odometer algorithm could perform high-precision positioning in a room-scale indoor environment, and the result was better than that of the monocular and wheel separately. By further combining complete SLAM system, ours almost have no bias.
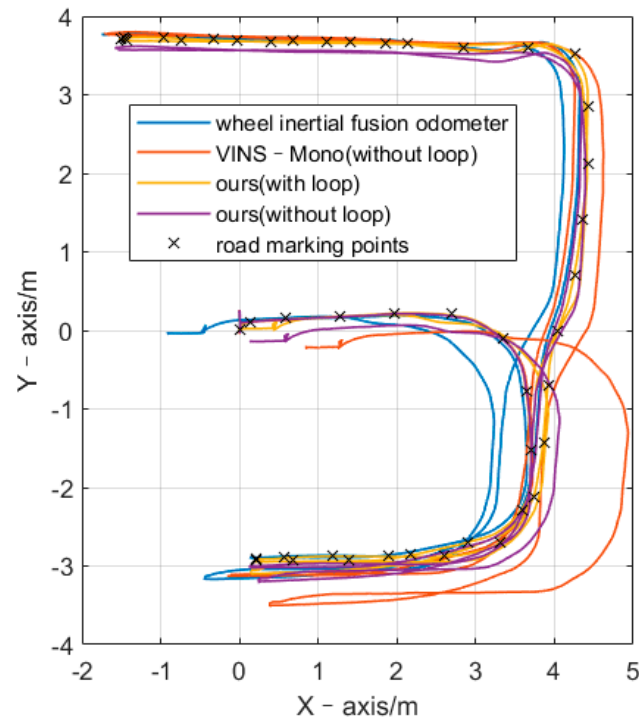
Figure 4. Path of pose measurement at room scale.

3.1.2. Floor Scale Experiment

Experimental conditions included:

① On the first floor of the building where the laboratory was located, the floor area was approximately 250×100 m;

② The robot ran in the central hall with 1 m/s (the area of the hall was approximately 15×15 m);

③ A large number of fixed marking points are arranged inside the corridor to obtain the real position of the robot in order to analyze the positioning error of the robot.

Anomalies during the experiment included:

① There was a considerable amount of dust on the ground, which decreased the friction between the wheels, leading to a slight slip during rapid turns and a significant slip during left-to-right translation;

② Due to the fast movement of the robot, the picture of the rolling shutter camera continued to exhibit disturbances;

③ There were cable manhole covers in numerous places in the corridor, the ground surface was uneven, and there were 2–3 cm step-like undulations. The robot vibrated significantly while passing over these obstacles;

④ The corridor contained semi-open areas and closed areas, the environment brightness changed significantly, and some areas were almost completely dark;

⑤ The walls around the robot in some areas were covered with tiles and had reflections;

⑥ The corridor and hall scenes had a high degree of similarity, lacking special landmarks. Loop detection was successfully performed only when the robot passed the hall halfway and finally returned to the hall;

⑦   During the experiment, pedestrians appeared several times.

The RViz interface at the end of test data playback is shown in Figure 5, which includes the raw picture, feature points, matched feature points, and the path.
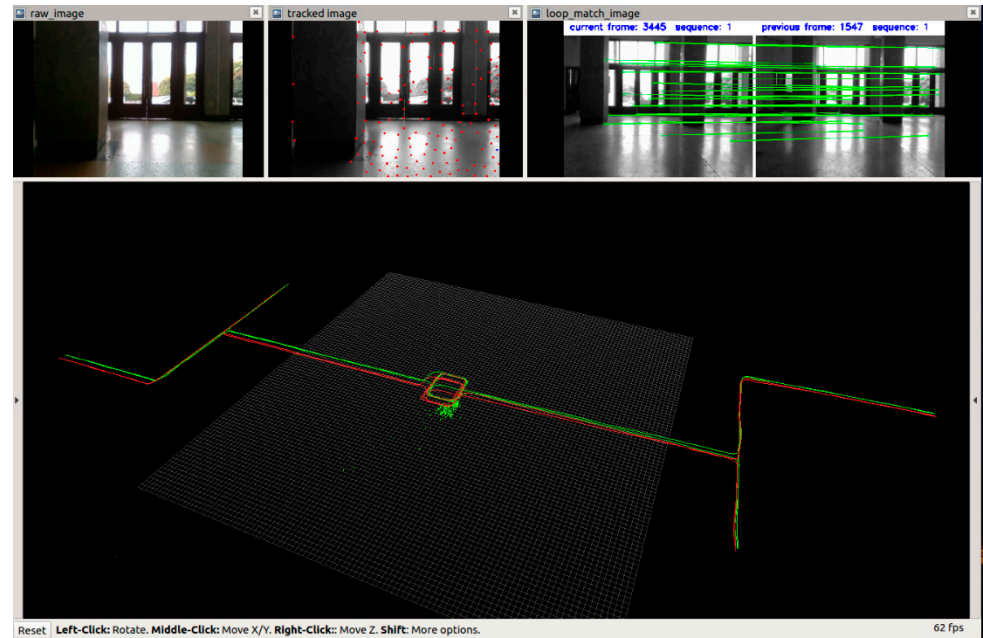

Figure 5. RViz interface at the end of test data playback.

Throughout the experiment, the car traveled for 889.2 s, the average speed was 0.87 m/s, the maximum speed was 1.36 m/s, the cumulative translation was 809.27 m, the cumulative rotation was 15,502.1°, and the chassis abnormal time was 26.983 s. As shown in Figure 6 and Table 2, the uneven ground made the wheel slip, so the error was very high. As shown in the floor-scale scenario, both the VINS-Mono algorithm with loop detection disabled and ours performed better; however, there was still a cumulative positioning error that could not be neglected. Compared with the complete SLAM system, VINS-Mono and ours both significantly improved the accuracy. Additionally, ours was better than that of VINS-Mono.

Table 2. Floor-scale pose measurements results.

Algorithm	ΔX (m)	ΔY (m)	ΔP (m)	ΔP (%)	ΔA (°)
wheel	−83.768	0.295	83.768	10.31	32.868
wheel inertial	−52.108	−0.826	52.115	6.42	6.404
VINS-Mono	10.692	−31.422	33.191	4.09	4.080
VINS-Mono (loop)	0.138	−5.281	5.281	0.65	3.086
ours	−2.025	0.947	2.235	0.28	3.661
ours (loop)	−0.295	0.165	0.338	0.04	3.172

3.2. Robustness Verification Experiment

This time, we created experiments with sensor measurement errors or even loss. Experimental conditions included intentionally blocking the camera, resulting in the visual signal being lost for approximately 15 s. This was undertaken in the laboratory, during the movement of the robot, during which the robot kept moving and turning.
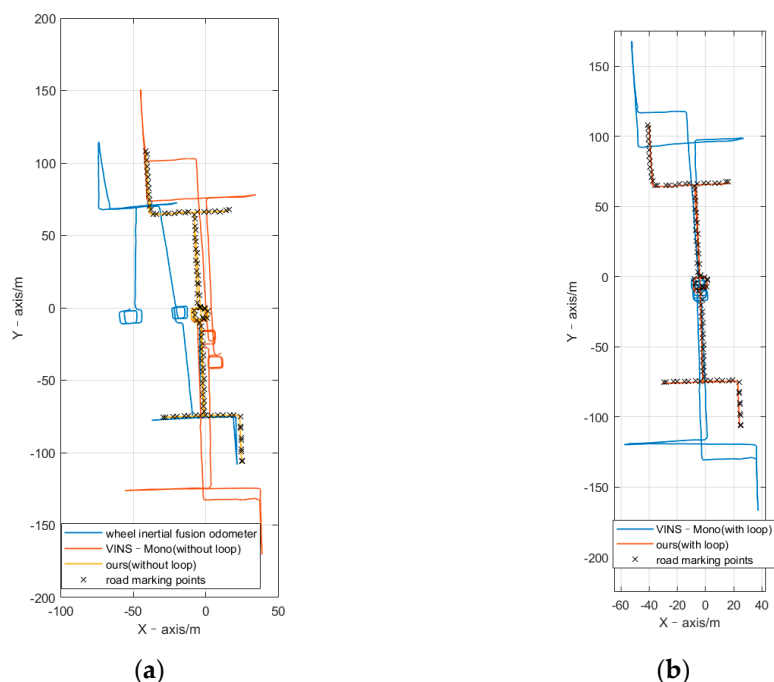
Figure 6. (**a**) Path of floor-scale without loop and (**b**) path of floor-scale with loop.

Throughout the experiment, the car traveled for 104.4 s, the average speed was 0.142 m/s, the maximum speed was 0.803 m/s, the cumulative translation was 17.809 m, the cumulative rotation was 1110.32°, and the chassis had no abnormalities. In the process of visual measurement loss, upon which the VINS-Mono is based, the state estimator was downgraded to inertial navigation dead reckoning. The error increased rapidly, thereby reducing the final positioning accuracy. As shown in Figure 7 and Table 3 during the loss of visual features, the path of the SLAM algorithm was the same as that measured by the wheel odometer. This shows that, although the positioning accuracy was affected by the loss of visual measurement, the pre-integration constraint can still provide absolute speed measurement; thus, the final positioning error was less than that of VINS-Mono.
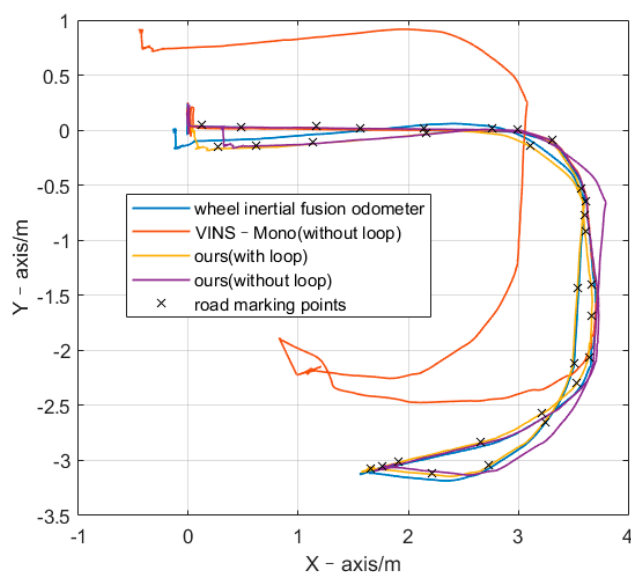

Figure 7. Path while visual tracking is failed.

Table 3. Pose measurement results when visual tracking was lost.

Algorithm	ΔX (m)	ΔY (m)	ΔP (m)	ΔP (%)	ΔA (°)
wheel	−0.042	0.106	0.114	0.64	−1.078
wheel inertial	−0.137	−0.014	0.137	0.77	−2.811
VINS-Mono	−0.442	0.890	0.994	5.58	−2.057
VINS-Mono (loop)	−0.030	0.015	0.025	0.19	−1.599
ours	0.295	0.011	0.295	1.66	−2.883
ours (loop)	0.020	−0.014	0.023	0.14	−1.718

4. Discussion

Aiming to solve the problems of low accuracy and poor robustness of the visual inertial SLAM algorithm, we designed and implemented a tightly coupled monocular visual inertial pose estimation algorithm that integrates wheel speed information. The state estimator, based on tightly coupled multi-sensor information, is used as the core of the algorithm, and the wheel odometer pre-integration that integrates wheel speed and IMU angular velocity avoids repeated integration in the iterative optimization process. The pose of the robot in the sliding window and the visual feature points are the state to be estimated. Then, the state is initialized based on the wheel odometer and IMU, so that the initial value of the state can be calculated quickly and reliably in both stationary and moving states. Finally, the residual constraint is added to the state estimator, and the optimal robot pose is solved through nonlinear optimization. Compared with the existing SLAM algorithm for many experiments, the SLAM algorithm in this paper can achieve higher pose accuracy and robustness in environments with more extreme situations. Experimental comparison results show that the SLAM algorithm proposed in this paper is feasible and effective.

In future research, our algorithm can be improved in the following ways. First, a motion capture system or D-GPS can be employed to obtain ground truth for quantitative pose comparison analysis. Second multiple monocular cameras with non-common view relationships can be used, which can improve robustness. When some cameras are blocked, the other cameras can still provide reliable visual measurement. Further, based on the robust robot pose measurement, using binocular or depth cameras to build a dense map of the environment can meet the requirements of robot navigation and obstacle avoidance.

Author Contributions: Conceptualization, G.P.; writing—original draft, Z.L.; writing—review and editing, J.P.; validation, D.H.; methodology, X.L.; resources, B.H. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the data may involve confidential information of our research institution.

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

Abbreviations

Symbols	Meaning
$\mathcal{X}$	variables to be estimated
i, j, k	key frame
$\mathcal{K}$	set of key frames
l	feature point
$\mathcal{L}$	set of feature points
x_k	state under key frame k
λ_l	inverse depth of feature point l
p_B^W	position(displacement coordinate of B relative to W)
q_B^W	attitude(quaternion form)
v_B^W	speed
b_a	accelerometer bias
b_g	gyroscope bias
$\hat{m}_{odomk}$	wheel odometer under key frame k
$\hat{m}$	means term m is the observation
$\triangleq$	defined as
$\{p_{xk} p_{yk}\}$	coordinate of the wheel odometer under key frame k
θ_k	angle of rotation of the wheel odometer under key frame k
$\mathcal{Z}$	observation
$\mathcal{Z}_{C_i}$	visual feature point observation under key frame i
$\hat{z}_{il}$	feature points observed under key frame i
$\mathcal{B}_{ij}$	IMU pre-integration between key frame i and key frame j
t	time
$\hat{a}_t$	pre-integration of accelerated speed at time t
$\hat{\omega}_t$	pre-integration of angular speed at time t
$\mathcal{O}_{ij}$	pre-fusion wheel odometer between key frame i and key frame j
$\mathcal{X}^*$	maximum posterior estimation of variables
$p(\mathcal{Z}\vert\mathcal{X})$	conditional probability of observing the occurrence of Z under the existed state X
$p(\mathcal{X})$	prior probability (edge probability)
$\|r(\cdot)\|_\Sigma^2$	Mahalanobis distance
Σ	covariance matrix
$\{r_p, H_p\}$	prior information from marginalization
r_C	visual residual
$r_\mathcal{O}$	wheel odometer residual
$r_\mathcal{B}$	IMU residual
$\rho(\cdot)$	Huber loss function
(p_{C_i}, q_{C_i})	spatial pose under key frame i
$r_{C_{jl}}$	error of the feature point l in the position of key frame j
$R_{\mathcal{F}_2}^{\mathcal{F}_1}$	rotation matrix from $\mathcal{F}_2$ coordinate system to $\mathcal{F}_1$ coordinate system
$\hat{\mathcal{P}}_l^{c_j}$	position where the feature point l is projected onto the unit ball in the key frame j
π_c^{-1}	back projection function
C_j	camera coordinate system of the key frame j
$\hat{\mathcal{P}}_l^{C_j}$	position projected on the unit ball in the key frame j
$[b_1 b_2]$	two orthogonal base vectors on the tangent plane of the unit ball and the projection lines with the feature points in the orthogonal direction
$\hat{\alpha}_{B_k+1}^{B_k}, \hat{\beta}_{B_k+1}^{B_k}, \hat{q}_{B_k+1}^{B_k}$	pre-integrated IMU measurement terms, observationα is position term; β is velocity term; q is rotation term
δ	represents the error term
$\delta\theta_{B_k+1}^{B_k}$	three-dimensional small perturbation
$p_{O_i}^{O_k}, \hat{p}_{O_i}^{O_k}$	displacement between key frame k and key frame i(the term with "ˆ" is the observation, without "ˆ" is the estimate)
η_{os}	position measurement noise, Gaussian
$D_{O_{k+1}}^{O_k}$	noise variance coefficient between key frame k and key frame k+1

η_g	angular velocity white noise
η_{b_g}	angular velocity zero error
$(\cdot)$	3×3 anti-symmetric matrix of Lie algebra
$J_{\delta x_i}^{\delta x_{i+1}}$	Jacobian matrix of the error value between the two frames before and after
$J_{\hat{x}_i}^{\hat{x}_{i+1}}$	Jacobian matrix of the nominal
R_O^B	known constant
$\hat{z}_{B_{k+1}}^{B_k}$	IMU pre-integration measurement between the key frames B_k and B_{k+1}
$x_{B_{k+1}}^{B_k}$	variable to be estimated related to the key frames B_k and B_{k+1}
$H_{B_{k+1}}^{B_k}$	constraint between the measurement $\hat{z}_{B_{k+1}}^{B_k}$ and the $x_{B_{k+1}}^{B_k}$
S_x, S_y	X-axis and Y-axis scale factors of the wheel odometer

References

1. Mur-Artal, R.; Montiel, J.M.; Tardos, J.D. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Trans. Robot.* **2015**, *31*, 1147–1163. [CrossRef]
2. Engel, J.; Schöps, T.; Cremers, D. LSD-SLAM: Large-scale direct monocular SLAM. In *European Conference on Computer Vision*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer: Zurich, Switzerland, 2014; pp. 834–849.
3. Engel, J.; Koltun, V.; Cremers, D. Direct sparse odometry. *IEEE Trans. Pattern Anal. Mach Intell.* **2018**, *40*, 611–625. [CrossRef] [PubMed]
4. Forster, C.; Pizzoli, M.; Scaramuzza, D. SVO: Fast semi-direct monocular visual odometry. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation, Hong Kong, China, 31 May–7 June 2014; pp. 15–22.
5. Valente, M.; Joly, C.; de la Fortelle, A. Evidential SLAM fusing 2D laser scanner and stereo camera. *Unmanned Syst.* **2019**, *7*, 149–159. [CrossRef]
6. Unsal, D.; Demirbas, K. Estimation of deterministic and stochastic IMU error parameters. In Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium, Myrtle Beach, SC, USA, 23–26 April 2012; pp. 862–868.
7. Wu, K.J.; Guo, C.X.; Georgiou, G.; Roumeliotis, S.I. VINS on wheels. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation, Marina Bay Sands, Singapore, 29 May–3 June 2017; pp. 5155–5162.
8. Karamat, T.B.; Lins, R.G.; Givigi, S.N.; Noureldin, A. Novel EKF-Based Vision/Inertial System Integration for Improved Navigation. *IEEE Trans. Instrum. Meas.* **2017**, *67*, 116–125. [CrossRef]
9. Qian, J.; Zi, B.; Wang, D.; Ma, Y.; Zhang, D. The design and development of an omni-directional mobile robot oriented to an intelligent manufacturing system. *Sensors* **2017**, *17*, 2073. [CrossRef] [PubMed]
10. Zhang, C.; Chen, L.; Yuan, S. ST-VIO: Visual Inertial Odometry Combined with Image Segmentation and Tracking. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 8562–8570. [CrossRef]
11. Qin, T.; Li, P.; Shen, S. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE Trans. Robot.* **2018**, *34*, 1004–1020. [CrossRef]
12. Qin, T.; Cao, S.; Pan, J.; Shen, S. A general optimization-based framework for global pose estimation with multiple sensors. *arXiv* **2019**, arXiv:1901.03642.
13. Qin, T.; Pan, J.; Cao, S.; Shen, S. A general optimization-based framework for local odometry estimation with multiple sensors. *arXiv* **2019**, arXiv:1901.03638.
14. Wang, J.; Shi, Z.; Zhong, Y. Visual SLAM incorporating wheel odometer for indoor robots. In Proceedings of the 2017 36th Chinese Control Conference (CCC), Dalian, China, 26–28 July 2017; pp. 5167–5172.
15. Zheng, F.; Tang, H.; Liu, Y.-H. Odometry-Vision-Based Ground Vehicle Motion Estimation with SE(2)-Constrained SE(3) Poses. *IEEE Trans. Cybern.* **2018**, *49*, 2652–2663. [CrossRef] [PubMed]
16. Yuan, W.; Li, Z.; Su, C.-Y. Multisensor-Based Navigation and Control of a Mobile Service Robot. *IEEE Trans. Syst. Man. Cybern. Syst.* **2019**, *49*, 1–11. [CrossRef]
17. Galvez-López, D.; Tardos, J. Bags of Binary Words for Fast Place Recognition in Image Sequences. *IEEE Trans. Robot.* **2012**, *28*, 1188–1197. [CrossRef]
18. Huber, P.J. *Robust Estimation of a Location Parameter*; Springer: New York, NY, USA, 1992.
19. Forster, C.; Carlone, L.; Dellaert, F.; Scaramuzza, D. IMU preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation. In *2015 Robotics Science and Systems*; Georgia Institute of Technology: Rome, Italy, 2015; pp. 1–20.