Monocular Visual-Inertial SLAM with Camera-IMU Extrinsic Automatic Calibration and Online Estimation

Linhao Pan(✉), Fuqing Tian, Wenjian Ying, and Bo She

Department of Weapon, Naval University of Engineering, Wuhan 430033, Hubei, China
jaypancool@gmail.com

Abstract. An approach of automatic calibration and online estimation for camera-IMU extrinsic parameters in monocular visual-inertial SLAM (Simultaneous Localization and Mapping) is proposed in this paper. Firstly, the camera-IMU extrinsic rotation is estimated with the hand-eye calibration as well as the gyroscope bias. Secondly, the scale factor, gravity and camera-IMU extrinsic translation are approximated without considering the accelerometer bias. All these parameters are refined with the gravitational magnitude and accelerometer bias taken into account at last. Furthermore, the camera-IMU extrinsic parameters are put into state vectors for online estimation. Experiment result with the EuRoC dataset shows that the algorithm automatically calibrates and estimates the camera-IMU extrinsic parameter with the extrinsic orientation and translation's error within 0.5° and 0.02 m separately, which contributes to the rapid use and accuracy of the VI-SLAM system.

Keywords: VI-SLAM · Sensor fusion · Initialization · Extrinsic calibration · State estimation

1 Introduction

VI-SLAM (Visual-Inertial Simultaneous Localization and Mapping) plays an important role in giving autonomous robots the ability to build the map of surroundings as well as estimate their states. Visual camera and inertial measurement unit (IMU) are ideal choice for SLAM since they could complement each other. On the one hand, the rich representation of environments projected by a camera helps to build a map and to estimate the trajectory of the robot up-to-scale. On the other hand, gyroscope and accelerometer of an IMU can obtain the angular velocity and linear acceleration of the sensor suite, which helps to recover the absolute scale information as well as make the gravity and the pitch and roll angle of the robot observable; however, the collected data will be affected by the measurement noise and drift with time [1]. Their superior size, weight and energy consumption make them widely used in the fields of robot navigation [2], UAVs [3] etc.

Recent years, several visual-inertial techniques have been presented in the field, such as the EKF based VI-SLAM [4, 5] algorithm and the nonlinear optimization

H. Yu et al. (Eds.): ICIRA 2019, LNAI 11743, pp. 706–721, 2019.
https://doi.org/10.1007/978-3-030-27538-9_61

methods [6, 7]. However, all the VI-SLAM algorithms depend heavily on accurate system initialization and prior precise extrinsic calibration of the 6DoF (Degree-of-Freedom) transformation between the camera and the IMU. Extrinsic parameters play a bridge role in the state transformation between camera reference frame and IMU reference frame.

At present, there are two main calibration methods for monocular camera-IMU extrinsic parameters: offline method and automatic calibration in system initialization. Offline calibration method requires technicians to carefully move the calibration checkboard in front of the sensor suite [8], which is complex and time-consuming. Automatic calibration in system initialization jointly estimate initial values and extrinsic parameters. Li [9], incorporating the camera-IMU transformation into the state vector, uses the extended Kalman filter (EKF) to estimate them. The convergence of the algorithm depends on the accuracy of the state estimation in initialization, and there is no systematic analysis of the results in the literature. Dong-Si proposed a geometric method to calibrate the camera-IMU extrinsic parameters in [10]. However, this method does not consider the noise of the sensor and tracking accuracy of the system will be affected by the accumulation of IMU bias. Yang and Shen [11], based on Lupton [12] and Martineli [13], calibrate the parameters (except for IMU bias) with an optimization-based linear estimator. The IMU bias is estimated as a state variable in the sliding window nonlinear estimator in their subsequent work of VI-SLAM system [14]. In the work of Huang [15], based on the work of Mur-Artal [16], a linear equation system is established to estimate the camera-IMU extrinsic parameters and other initialization parameters. This method has high initialization accuracy, but the camera-IMU extrinsic parameters become fixed after initialization without online estimation.

In this paper, we realize a VI-SLAM algorithm with monocular camera-IMU extrinsic automatic calibration and online estimation. Without knowing the mechanical configuration of the sensor suite, the scale factor, gravity, IMU biases and extrinsic parameters are jointly estimated in the initialization, as well as online estimation of camera-IMU extrinsic parameters during motion.

The rest of this paper is organized as follows. Section 2 describes the preliminaries of this algorithm. Then the initialization process with camera-IMU extrinsic automatic calibration is proposed in Sect. 3. Section 4 describes online estimation algorithm of camera-IMU extrinsic parameters. Experimental results are shown in Sect. 5. Finally, conclusions are given in Sect. 6.

2 Preliminaries

This section provides the necessary explanations for the notation and geometric concepts involved in this article. In addition, the relationship between reference frames and the IMU preintegration model on manifold are also described.

2.1 Notation

The matrices and vectors used here are indicated in bold uppercase and lowercase respectively. The letter in the upper right corner of the vector indicates the reference

frames of the vector, e.g. $\boldsymbol{v}^{\boldsymbol{W}}$ for the vector $\boldsymbol{v}$ expressed in frame W. Incorporating geometric meaning, $\boldsymbol{p}_{\boldsymbol{B}}^{\boldsymbol{C}}$ and $\boldsymbol{v}_{\boldsymbol{B}}^{\boldsymbol{W}}$ represent the point $\boldsymbol{p}_{\boldsymbol{B}}$ and velocity vector $\boldsymbol{v}_{\boldsymbol{B}}$ in the reference frames C and W respectively. In the frame C, the rotation matrix and translation matrix of the frame B are represented by $\boldsymbol{R}_{\boldsymbol{CB}}$ and $\boldsymbol{T}_{\boldsymbol{CB}}$ respectively.

2.2 Reference Frames

The VI-SLAM system mainly involves four frames: camera frame C, IMU body frame B, world frame W and inertial frame E. As shown in Fig. 1,

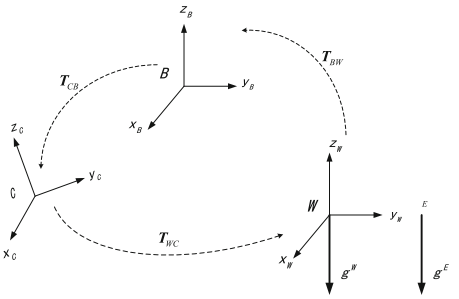

Fig. 1. The Reference frame transformation.

The monocular camera and the IMU are fixed by external devices, and the transformation matrix $\boldsymbol{T}_{CB} = \left\{\boldsymbol{R}_{CB} \middle| \boldsymbol{p}_B^C\right\}$ between them needs to be calibrated. Since the VI-SLAM system measures the relative motion, the absolute attitude in the earth's inertial frame E cannot be determined. Therefore, the coordinate system of the first keyframe determined by the VI-SLAM system generally coincides with the world frame W. The goal is to calibrate the rotation matrix $\boldsymbol{R}_{\boldsymbol{CB}} \in SO(3)$ and translation vector $\boldsymbol{p}_{\boldsymbol{B}}^{\boldsymbol{C}} \in \mathbb{R}^3$ between camera frame C and IMU body frame B and estimate the gravitational acceleration in the world frame W. By aligning the gravity in the system's world frame $\boldsymbol{g}^{\boldsymbol{W}}$ with the gravity in the earth's inertial frame $\boldsymbol{g}^{\boldsymbol{E}}$, the absolute pose of the system in the inertial frame is determined. Considering the scale factor s, the transformation between camera frame C and IMU body B frame is

$$\boldsymbol{R}_{\boldsymbol{WB}} = \boldsymbol{R}_{\boldsymbol{WC}} \cdot \boldsymbol{R}_{\boldsymbol{CB}} \tag{1}$$

$$\boldsymbol{p}_{\boldsymbol{B}}^{\boldsymbol{W}} = s \cdot \boldsymbol{p}_{\boldsymbol{C}}^{\boldsymbol{W}} + \boldsymbol{R}_{\boldsymbol{WC}} \cdot \boldsymbol{p}_{\boldsymbol{B}}^{\boldsymbol{C}} \tag{2}$$

2.3 Preintegration

The IMU sensor acquires angular velocity and acceleration of the sensor w.r.t. the IMU body frame B at a certain frequency. However, the gyroscope and accelerometer are subject to white sensor noises $\boldsymbol{\eta}_{\mathbf{a}}$ and $\boldsymbol{\eta}_{\mathbf{g}}$, as well as low-frequency drift biases $\boldsymbol{b}_{\boldsymbol{a}}$ and

$\boldsymbol{b}_g$. In the case where the initial state of the system is known, the state estimation of the system can be propagated by integrating the IMU measurements. However, this method is based on the initial state. When the nonlinear optimization adjusts system initial state, the integration process needs to be repeated.

To avoid repeated integration, the concept of preintegraed IMU measurement on the manifold space was proposed by Forster et al [17]. Assuming two consecutive keyframes at time i and j, and the IMU measurement value is constant during the sampling interval. So the pose and velocity relationship can be computed by numerical integration of all measurements within this period

$$\boldsymbol{R}_{WB_j} = \boldsymbol{R}_{WB_i} \prod_{k=i}^{j-1} Exp\left(\left(\boldsymbol{\omega}^{B_k} - \boldsymbol{b}_g^k - \boldsymbol{\eta}_g^k\right)\Delta t\right)$$

$$\boldsymbol{v}_{B_j}^W = \boldsymbol{v}_{B_i}^W + \boldsymbol{g}^W \Delta t_{ij} + \sum_{k=i}^{j-1} \boldsymbol{R}_{WB_k}\left(\boldsymbol{a}^{B_k} - \boldsymbol{b}_a^k - \boldsymbol{\eta}_a^k\right)\Delta t \tag{3}$$

$$\boldsymbol{p}_{B_j}^W = \boldsymbol{p}_{B_i}^W + \sum_{k=i}^{j-1}\left(\boldsymbol{v}_{B_k}^W \Delta t + \frac{1}{2}\boldsymbol{g}^W \Delta t^2 + \frac{1}{2}\boldsymbol{R}_{WB_k}\left(\boldsymbol{a}^{B_k} - \boldsymbol{b}_a^k - \boldsymbol{\eta}_a^k\right)\Delta t^2\right)$$

Where Δt denotes the IMU sampling interval, $\Delta t_{ij} = \sum_{i}^{j-1} \Delta t$ represents time interval between two consecutive frames. According to the definition in [17], $Exp(\cdot)$ maps Lie algebra $so(3)$ to Lie group $SO(3)$. When the bias is assumed to remain constant between two image acquisition moments, a small bias correction w.r.t. previously estimated $\overline{\boldsymbol{b}}_{(\cdot)}^{\boldsymbol{i}}$ could be $\delta \boldsymbol{b}_{(\cdot)}^{\boldsymbol{i}}$. The Eq. (8) can be rewritten as

$$\boldsymbol{R}_{WB_j} = \boldsymbol{R}_{WB_i} \Delta \overline{\boldsymbol{R}}_{B_i B_j} Exp\left(\boldsymbol{J}_{\Delta \overline{\boldsymbol{R}}_{B_i B_j}}^{\boldsymbol{g}} \cdot \delta \boldsymbol{b}_g^{\boldsymbol{i}}\right)$$

$$\boldsymbol{v}_{B_j}^W = \boldsymbol{v}_{B_i}^W + \boldsymbol{g}^W \Delta t_{ij} + \boldsymbol{R}_{WB_i}\left(\Delta \overline{\boldsymbol{v}}_{B_j}^{B_i} + \boldsymbol{J}_{\Delta \overline{\boldsymbol{v}}_{B_j}^{B_i}}^{\boldsymbol{g}} \cdot \delta \boldsymbol{b}_g^{\boldsymbol{i}} + \boldsymbol{J}_{\Delta \overline{\boldsymbol{v}}_{B_j}^{B_i}}^{\boldsymbol{a}} \cdot \delta \boldsymbol{b}_a^{\boldsymbol{i}}\right) \tag{4}$$

$$\boldsymbol{p}_{B_j}^W = \boldsymbol{p}_{B_i}^W + \boldsymbol{v}_{B_i}^W \Delta t_{ij} + \frac{1}{2}\boldsymbol{g}^W \Delta t_{ij}^2 + \boldsymbol{R}_{WB_i}\left(\Delta \overline{\boldsymbol{p}}_{B_j}^{B_i} + \boldsymbol{J}_{\Delta \overline{\boldsymbol{p}}_{B_j}^{B_i}}^{\boldsymbol{g}} \cdot \delta \boldsymbol{b}_g^{\boldsymbol{i}} + \boldsymbol{J}_{\Delta \overline{\boldsymbol{p}}_{B_j}^{B_i}}^{\boldsymbol{a}} \cdot \delta \boldsymbol{b}_a^{\boldsymbol{i}}\right)$$

$\boldsymbol{J}_{(\cdot)}^{\boldsymbol{g}}$ and $\boldsymbol{J}_{(\cdot)}^{\boldsymbol{a}}$ are Jacobian matrix of preintegration measurements relative to bias estimation, which is deduced in the appendix of paper [17]. $\Delta \overline{\boldsymbol{R}}_{B_i B_j}$, $\Delta \overline{\boldsymbol{v}}_{B_j}^{B_i}$ and $\Delta \overline{\boldsymbol{p}}_{B_j}^{B_i}$ are the terms of preintegration which are independent of the states at time i and the gravity, used to describe the relative motion of two frames

$$\Delta\overline{\boldsymbol{R}}_{\boldsymbol{B}_i\boldsymbol{B}_j} = \prod_{k=i}^{j-1} Exp\left(\left(\boldsymbol{\omega}^{\boldsymbol{B}_k} - \overline{\boldsymbol{b}}_{\boldsymbol{g}}^{\boldsymbol{i}}\right)\Delta t\right)$$

$$\Delta\overline{\boldsymbol{v}}_{\boldsymbol{B}_j}^{\boldsymbol{B}_i} = \sum_{k=i}^{j-1} \Delta\overline{\boldsymbol{R}}_{\boldsymbol{B}_i\boldsymbol{B}_k}\left(\boldsymbol{a}^{\boldsymbol{B}_k} - \overline{\boldsymbol{b}}_{\boldsymbol{a}}^{\boldsymbol{i}}\right)\Delta t \tag{5}$$

$$\Delta\overline{\boldsymbol{p}}_{\boldsymbol{B}_j}^{\boldsymbol{B}_i} = \sum_{k=i}^{j-1}\left(\Delta\overline{\boldsymbol{v}}_{\boldsymbol{B}_k}^{\boldsymbol{B}_i}\Delta t + \frac{1}{2}\Delta\overline{\boldsymbol{R}}_{\boldsymbol{B}_i\boldsymbol{B}_k}\left(\boldsymbol{a}^{\boldsymbol{B}_k} - \overline{\boldsymbol{b}}_{\boldsymbol{a}}^{\boldsymbol{i}}\right)\Delta t^2\right)$$

3 Initialization with Camera-IMU Extrinsic Automatic Calibration

This section elaborates on the initialization method with camera-IMU extrinsic automatic calibration. This method jointly calibrates the camera-IMU extrinsic parameters $\boldsymbol{T}_{\boldsymbol{CB}}$, as well as estimates factor scale s, gravity acceleration in the world frame $\boldsymbol{g}^{\boldsymbol{W}}$, biases of gyroscope and accelerometer $\boldsymbol{b}_{\boldsymbol{a}}$ and $\boldsymbol{b}_{\boldsymbol{g}}$.

3.1 Camera-IMU Extrinsic Orientation Calibration

The extrinsic rotation between the monocular camera and the IMU is very important for the robustness of the VI-SLAM system. Excessive deviation can cause the system initialization to collapse. The hand-eye calibration method is used to align the rotations of the camera with the integrated IMU rotations. Because the monocular camera can track the pose of the system, to detect the relative rotation $\boldsymbol{R}_{\boldsymbol{C}_i\boldsymbol{C}_{i+1}}$ between consecutive frames. In addition, the angular velocity measured by the gyroscope can be integrated to obtain relative rotation $\boldsymbol{R}_{\boldsymbol{B}_i\boldsymbol{B}_{i+1}}$ in the IMU body frame. So it leads to

$$\boldsymbol{R}_{\boldsymbol{B}_i\boldsymbol{B}_{i+1}} \cdot \boldsymbol{R}_{\boldsymbol{BC}} = \boldsymbol{R}_{\boldsymbol{BC}} \cdot \boldsymbol{R}_{\boldsymbol{C}_i\boldsymbol{C}_{i+1}} \tag{6}$$

With the quaternion representation, (6) can be described as

$$\begin{aligned} &\boldsymbol{q}_{\boldsymbol{B}_i\boldsymbol{B}_{i+1}} \otimes \boldsymbol{q}_{\boldsymbol{BC}} = \boldsymbol{q}_{\boldsymbol{BC}} \otimes \boldsymbol{q}_{\boldsymbol{C}_i\boldsymbol{C}_{i+1}} \\ &\Rightarrow \left[\left[\boldsymbol{q}_{\boldsymbol{B}_i\boldsymbol{B}_{i+1}}\right]_L - \left[\boldsymbol{q}_{\boldsymbol{C}_i\boldsymbol{C}_{i+1}}\right]_R\right]\boldsymbol{q}_{\boldsymbol{BC}} = \boldsymbol{Q}_{i,i+1} \cdot \boldsymbol{q}_{\boldsymbol{BC}} = \boldsymbol{0}_{\boldsymbol{4}\times\boldsymbol{1}} \end{aligned} \tag{7}$$

Linear over-determined equation can be established for temporally continuous frames

$$\begin{bmatrix} \alpha_{0,1} \cdot \boldsymbol{Q}_{0,1} \\ \alpha_{1,2} \cdot \boldsymbol{Q}_{1,2} \\ \vdots \\ \alpha_{N-1,N} \cdot \boldsymbol{Q}_{N-1,N} \end{bmatrix} \boldsymbol{q}_{BC} = \boldsymbol{Q}_N \cdot \boldsymbol{q}_{BC} = \boldsymbol{0} \tag{8}$$

N indicates the number of frames used when extrinsic rotation converges; $\alpha_{N-1,N}$ is a weight for outlier handling. As the extrinsic rotation calibration runs with incoming measurements, the previously estimated result $\widehat{\boldsymbol{R}}_{\boldsymbol{BC}}$ can be used as the initial value to weight the residual

$$r_{i,i+1} = \arccos\left(\left(tr\left(\widehat{\boldsymbol{R}}_{\boldsymbol{BC}}^{-1}\boldsymbol{R}_{\boldsymbol{B}_i\boldsymbol{B}_{i+1}}^{-1}\widehat{\boldsymbol{R}}_{\boldsymbol{BC}}\boldsymbol{R}_{\boldsymbol{C}_i\boldsymbol{C}_{i+1}}\right) - 1\right)/2\right) \tag{9}$$

The weight is a function of the residual

$$\alpha_{i,i+1} = \begin{cases} 1, & r_{i,i+1} < t_0 \\ \frac{t_0}{r_{i,i+1}}, & otherwise \end{cases} \tag{10}$$

t_0 is the threshold. The solution to (8) can be found as the right unit singular vector corresponding to the smallest singular value of $\boldsymbol{Q}_N$.

3.2 Gyroscope Bias Estimation

The gyroscope bias can be estimated by the rotation relationship of consecutive keyframes. This paper assumes that the gyroscope bias remains constant in the initialization stage, and the initial gyroscope bias $\overline{\boldsymbol{b}_g}$ is 0.

Substituting the extrinsic rotation matrix $\widehat{\boldsymbol{R}}_{\boldsymbol{BC}}$ estimated in Sect. 3.1 into (4)

$$\left(\boldsymbol{R}_{\boldsymbol{WC}_i} \cdot \widehat{\boldsymbol{R}}_{\boldsymbol{CB}}\right)^T \left(\boldsymbol{R}_{\boldsymbol{WC}_{i+1}} \cdot \widehat{\boldsymbol{R}}_{\boldsymbol{CB}}\right) = \Delta\overline{\boldsymbol{R}}_{\boldsymbol{B}_i\boldsymbol{B}_{i+1}} Exp\left(\boldsymbol{J}^{\boldsymbol{g}}_{\Delta\overline{\boldsymbol{R}}_{\boldsymbol{B}_i\boldsymbol{B}_{i+1}}} \cdot \delta\boldsymbol{b}_{\boldsymbol{g}}\right) \tag{11}$$

For all keyframes during initialization, we use the minimum function for bias estimation,

$$\delta\boldsymbol{b}_{\boldsymbol{g}}^* = \underset{\delta\boldsymbol{b}_{\boldsymbol{g}}}{\arg\min} \sum_{i=1}^{N-1} \left\| Log\left(\left(\Delta\overline{\boldsymbol{R}}_{\boldsymbol{B}_i\boldsymbol{B}_{i+1}} Exp\left(\boldsymbol{J}^{\boldsymbol{g}}_{\Delta\overline{\boldsymbol{R}}_{\boldsymbol{B}_i\boldsymbol{B}_{i+1}}} \cdot \delta\boldsymbol{b}_{\boldsymbol{g}}\right)\right)^T \widehat{\boldsymbol{R}}_{\boldsymbol{BC}}\boldsymbol{R}_{\boldsymbol{C}_i\boldsymbol{W}}\boldsymbol{R}_{\boldsymbol{WC}_{i+1}}\widehat{\boldsymbol{R}}_{\boldsymbol{CB}}\right)\right\|^2 \tag{12}$$

where $\|\cdot\|$ is the L2-norm, $Log(\cdot)$ is the inverse of $Exp(\cdot)$. $\boldsymbol{R}_{\boldsymbol{WC}_i}$ and $\Delta\overline{\boldsymbol{R}}_{\boldsymbol{B}_i\boldsymbol{B}_{i+1}}$ are known to be obtained by monocular camera pose tracking and preintegration of gyroscope measurements respectively. This equation can be solved with Gauss-Newton algorithm. The final estimated gyroscope bias is $\hat{\boldsymbol{b}}_{\boldsymbol{g}} = \overline{\boldsymbol{b}_{\boldsymbol{g}}} + \delta\boldsymbol{b}_{\boldsymbol{g}}^* = \delta\boldsymbol{b}_{\boldsymbol{g}}^*$.

3.3 Scale, Gravity and Translation Approximation Without Accelerometer Bias

Once the gyroscope bias has been estimated, the preintegrations can be rectified by Eq. (5). And continue to estimate the scale factor s, gravity $\boldsymbol{g}^{\boldsymbol{W}}$ and extrinsic translation $\boldsymbol{p}_{\boldsymbol{B}}^{\boldsymbol{C}}$ approximately. Since the gravity and accelerometer bias are hard to be distinguished, accelerometer bias is not considered in this stage. So the $\boldsymbol{J}^{a}_{\Delta\overline{\boldsymbol{v}}^{\boldsymbol{B}_i}_{\boldsymbol{B}_j}}$ and $\boldsymbol{J}^{a}_{\Delta\overline{\boldsymbol{p}}^{\boldsymbol{B}_i}_{\boldsymbol{B}_j}}$ can be set to zero. Substituting Eq. (2) into the third equation of Eq. (4), the relationship of two consecutive keyframes can be obtained,

$$\begin{aligned} s\cdot\boldsymbol{p}^{\boldsymbol{W}}_{\boldsymbol{C}_{i+1}} &= s\cdot\boldsymbol{p}^{\boldsymbol{W}}_{\boldsymbol{C}_i}+\boldsymbol{v}^{\boldsymbol{W}}_{\boldsymbol{B}_i}\cdot\Delta t_{i,i+1}+\frac{1}{2}\boldsymbol{g}^{\boldsymbol{W}}\cdot\Delta t^2_{i,i+1} \\ &\quad+\boldsymbol{R}_{\boldsymbol{WC}_i}\cdot\widehat{\boldsymbol{R}}_{\boldsymbol{CB}}\cdot\Delta\overline{\boldsymbol{p}}^{\boldsymbol{B}_i}_{\boldsymbol{B}_{i+1}}+\left(\boldsymbol{R}_{\boldsymbol{WC}_i}-\boldsymbol{R}_{\boldsymbol{WC}_{i+1}}\right)\boldsymbol{p}^{\boldsymbol{C}}_{\boldsymbol{B}} \end{aligned} \tag{13}$$

The goal is to estimate s, $\boldsymbol{g}^{\boldsymbol{W}}$ and $\boldsymbol{p}^{\boldsymbol{C}}_{\boldsymbol{B}}$. Using two relations between three consecutive keyframes to eliminate velocities $\boldsymbol{v}^{\boldsymbol{W}}_{\boldsymbol{B}_i}$, which leads to the following expression:

$$\begin{bmatrix}\boldsymbol{\lambda}(i) & \boldsymbol{\beta}(i) & \boldsymbol{\varphi}(i)\end{bmatrix}\begin{bmatrix}s\\ \boldsymbol{g}^{\boldsymbol{W}}\\ \boldsymbol{p}^{\boldsymbol{C}}_{\boldsymbol{B}}\end{bmatrix}=\boldsymbol{\gamma}(i) \tag{14}$$

Writing the subscript i, $i+1$, $i+2$ as 1, 2, 3, we have:

$$\begin{aligned} \boldsymbol{\lambda}(i) &= \left(\boldsymbol{p}^{\boldsymbol{W}}_{\boldsymbol{C}_3}-\boldsymbol{p}^{\boldsymbol{W}}_{\boldsymbol{C}_2}\right)\Delta t_{12}+\left(\boldsymbol{p}^{\boldsymbol{W}}_{\boldsymbol{C}_1}-\boldsymbol{p}^{\boldsymbol{W}}_{\boldsymbol{C}_2}\right)\Delta t_{23} \\ \boldsymbol{\beta}(i) &= -\frac{1}{2}\left(\Delta t^2_{12}\Delta t_{23}+\Delta t^2_{23}\Delta t_{12}\right)\mathbf{I}_{3\times3} \\ \boldsymbol{\varphi}(i) &= \left(\boldsymbol{R}_{\boldsymbol{WC}_3}-\boldsymbol{R}_{\boldsymbol{WC}_2}\right)\Delta t_{12}+\left(\boldsymbol{R}_{\boldsymbol{WC}_1}-\boldsymbol{R}_{\boldsymbol{WC}_2}\right)\Delta t_{23} \\ \boldsymbol{\gamma}(i) &= \boldsymbol{R}_{\boldsymbol{WC}_1}\cdot\widehat{\boldsymbol{R}}_{\boldsymbol{CB}}\cdot\Delta\overline{\boldsymbol{v}}^{\boldsymbol{B}_1}_{\boldsymbol{B}_2}\Delta t_{12}\Delta t_{23}+\boldsymbol{R}_{\boldsymbol{WC}_2}\cdot\widehat{\boldsymbol{R}}_{\boldsymbol{CB}}\cdot\Delta\overline{\boldsymbol{p}}^{\boldsymbol{B}_2}_{\boldsymbol{B}_3}\Delta t_{12} \\ &\quad-\boldsymbol{R}_{\boldsymbol{WC}_1}\cdot\widehat{\boldsymbol{R}}_{\boldsymbol{CB}}\cdot\Delta\overline{\boldsymbol{p}}^{\boldsymbol{B}_1}_{\boldsymbol{B}_2}\Delta t_{23} \end{aligned} \tag{15}$$

For N consecutive keyframes, a linear over-determined equation $\boldsymbol{A}_{3(N-2)\times7}\cdot\boldsymbol{x}_{7\times1}=\boldsymbol{B}_{3(N-2)\times1}$ can be stacked. $\hat{s}$, $\hat{\boldsymbol{g}}^{\boldsymbol{W}}$, $\hat{\boldsymbol{p}}^{\boldsymbol{C}}_{\boldsymbol{B}}$ can be solved by SVD decomposition. Note that there are $3(N-2)$ linear constraints and 7 unknowns, so at least 5 keyframes is required to solve the equation.

3.4 Accelerometer Bias Estimation, and Scale, Gravity and Translation Refinement

Assuming the gravity $\boldsymbol{g}^{\boldsymbol{E}}$ in the inertial frame E is known, and $G=9.8$ represents the magnitude of the gravitational acceleration as well as $\bar{\boldsymbol{g}}^{\boldsymbol{E}}=(0,0,-1)$ represents its

direction. It is stipulated that the earth's inertial frame E coincides with the origin of the world frame W, the rotation $\boldsymbol{R}_{\boldsymbol{WE}}$ between them can be computed as follows, shown in Fig. 2:

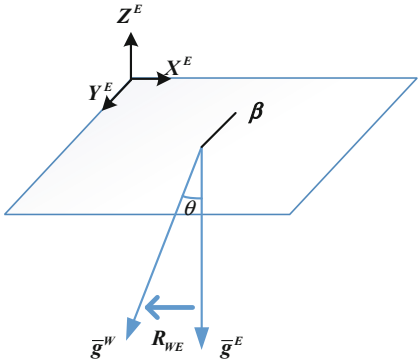

Fig. 2. Diagram of gravity acceleration direction angle.

$$\boldsymbol{R}_{\boldsymbol{WE}} = Exp(\theta\boldsymbol{\beta})$$

$$\boldsymbol{\beta} = \frac{\bar{\boldsymbol{g}}^{\boldsymbol{E}} \times \bar{\boldsymbol{g}}^{\boldsymbol{W}}}{\left\|\bar{\boldsymbol{g}}^{\boldsymbol{E}} \times \bar{\boldsymbol{g}}^{\boldsymbol{W}}\right\|}, \theta = \text{atan2}\left(\left\|\bar{\boldsymbol{g}}^{\boldsymbol{E}} \times \bar{\boldsymbol{g}}^{\boldsymbol{W}}\right\|, \bar{\boldsymbol{g}}^{\boldsymbol{E}} \cdot \bar{\boldsymbol{g}}^{\boldsymbol{W}}\right) \tag{16}$$

As $\bar{\boldsymbol{g}}^{\boldsymbol{W}} = \hat{\boldsymbol{g}}^{\boldsymbol{W}} / \left\|\hat{\boldsymbol{g}}^{\boldsymbol{W}}\right\|$ represents the gravity direction in the world frame W estimated in Sect. 3.3. This rotation can be optimized by appending perturbation $\delta\boldsymbol{\theta} \in \mathbb{R}^{3\times1}$:

$$\boldsymbol{g}^{\boldsymbol{W}} = \boldsymbol{R}_{\boldsymbol{WE}} \cdot Exp(\delta\boldsymbol{\theta})\boldsymbol{g}^{\boldsymbol{E}} \approx \boldsymbol{R}_{\boldsymbol{WE}} \cdot \boldsymbol{g}^{\boldsymbol{E}} - \boldsymbol{R}_{\boldsymbol{WE}}(\boldsymbol{g}^{\boldsymbol{E}})^{\wedge}\delta\boldsymbol{\theta}$$

$$\delta\boldsymbol{\theta} = \left[\delta\boldsymbol{\theta}_{\boldsymbol{xy}}^{T}, 0\right]^{T}, \delta\boldsymbol{\theta}_{\boldsymbol{xy}} = \left[\delta\boldsymbol{\theta}_{\boldsymbol{x}}, \delta\boldsymbol{\theta}_{\boldsymbol{y}}\right]^{T} \tag{17}$$

Substituting Eq. (17) into Eq. (13) and including the effect of accelerometer bias, we have:

$$\begin{aligned} s \cdot \boldsymbol{p}_{\boldsymbol{C}_{i+1}}^{\boldsymbol{W}} =& s \cdot \boldsymbol{p}_{\boldsymbol{C}_i}^{\boldsymbol{W}} + \boldsymbol{v}_{\boldsymbol{B}_i}^{\boldsymbol{W}} \Delta t_{i,i+1} - \frac{1}{2}\boldsymbol{R}_{\boldsymbol{WE}}(\boldsymbol{g}^{\boldsymbol{E}})^{\wedge}\delta\boldsymbol{\theta} \cdot \Delta t_{i,i+1}^{2} \\ &+ \boldsymbol{R}_{\boldsymbol{WC}_i} \cdot \widehat{\boldsymbol{R}}_{\boldsymbol{CB}}\left(\Delta\bar{\boldsymbol{p}}_{\boldsymbol{B}_{i+1}}^{\boldsymbol{B}_i} + \boldsymbol{J}_{\Delta\bar{\boldsymbol{p}}_{\boldsymbol{B}_{i+1}}^{\boldsymbol{B}_i}}^{a} \cdot \delta\boldsymbol{b}_a\right) + \left(\boldsymbol{R}_{\boldsymbol{WC}_i} - \boldsymbol{R}_{\boldsymbol{WC}_{i+1}}\right)\boldsymbol{p}_{\boldsymbol{B}}^{\boldsymbol{C}} \\ &+ \frac{1}{2}\boldsymbol{R}_{\boldsymbol{WE}} \cdot \boldsymbol{g}^{\boldsymbol{E}} \cdot \Delta t_{i,i+1}^{2} \end{aligned} \tag{18}$$

Considering the constraints between three consecutive keyframes as well, we can construct the following linear equations:

$$[\boldsymbol{\lambda}(i) \quad \boldsymbol{\alpha}(i) \quad \boldsymbol{\phi}(i) \quad \boldsymbol{\varphi}(i)] \begin{bmatrix} s \\ \delta\boldsymbol{\theta}_{xy} \\ \delta\boldsymbol{b}_a \\ \boldsymbol{p}_B^C \end{bmatrix} = \boldsymbol{\chi}(i) \tag{19}$$

Where $\boldsymbol{\lambda}(i)$, $\boldsymbol{\varphi}(i)$ remains the same as Eq. (15), $\boldsymbol{\alpha}(i), \boldsymbol{\phi}(i), \boldsymbol{\chi}(i)$ are computed as follow:

$$\begin{aligned}
\boldsymbol{\alpha}(i) &= \left[\frac{1}{2}\boldsymbol{R}_{\boldsymbol{WE}}(\boldsymbol{g}^{\boldsymbol{E}})^{\wedge}\left(\Delta t_{12}^2\Delta t_{23} + \Delta t_{23}^2\Delta t_{12}\right)\right]_{(:,1:2)} \\
\boldsymbol{\phi}(i) &= \boldsymbol{R}_{\boldsymbol{WC}_1} \cdot \widehat{\boldsymbol{R}}_{\boldsymbol{CB}} \cdot \boldsymbol{J}^{a}_{\Delta\overline{\boldsymbol{p}}^{\boldsymbol{B}_1}_{\boldsymbol{B}_2}} \Delta t_{23} - \boldsymbol{R}_{\boldsymbol{WC}_2} \cdot \widehat{\boldsymbol{R}}_{\boldsymbol{CB}} \cdot \boldsymbol{J}^{a}_{\Delta\overline{\boldsymbol{p}}^{\boldsymbol{B}_2}_{\boldsymbol{B}_3}} \Delta t_{12} \\
&\quad - \boldsymbol{R}_{\boldsymbol{WC}_1} \cdot \widehat{\boldsymbol{R}}_{\boldsymbol{CB}} \cdot \boldsymbol{J}^{a}_{\Delta\overline{\boldsymbol{v}}^{\boldsymbol{B}_1}_{\boldsymbol{B}_2}} \Delta t_{12}\Delta t_{23} \\
\boldsymbol{\chi}(i) &= \boldsymbol{R}_{\boldsymbol{WC}_2} \cdot \widehat{\boldsymbol{R}}_{\boldsymbol{CB}} \cdot \Delta\overline{\boldsymbol{p}}^{\boldsymbol{B}_2}_{\boldsymbol{B}_3}\Delta t_{12} + \boldsymbol{R}_{\boldsymbol{WC}_1} \cdot \widehat{\boldsymbol{R}}_{\boldsymbol{CB}} \cdot \Delta\overline{\boldsymbol{v}}^{\boldsymbol{B}_1}_{\boldsymbol{B}_2}\Delta t_{12}\Delta t_{23} \\
&\quad + \frac{1}{2}\boldsymbol{R}_{\boldsymbol{WE}} \cdot \boldsymbol{g}^{\boldsymbol{E}}\left(\Delta t_{12}^2\Delta t_{23} + \Delta t_{12}\Delta t_{23}^2\right) - \boldsymbol{R}_{\boldsymbol{WC}_1} \cdot \widehat{\boldsymbol{R}}_{\boldsymbol{CB}} \cdot \Delta\overline{\boldsymbol{p}}^{\boldsymbol{B}_1}_{\boldsymbol{B}_2}\Delta t_{23}
\end{aligned} \tag{20}$$

$[\cdot]_{(:,1:2)}$ in the $\boldsymbol{\alpha}(i)$ means the first two columns of the matrix.

Similar to above, a linear over-determined equation $\boldsymbol{A}_{3(N-2)\times 9} \cdot \boldsymbol{x}_{9\times 1} = \boldsymbol{B}_{3(N-2)\times 1}$ can be constructed to calculate the $\delta\boldsymbol{b}_a^*$, s^*, $\delta\boldsymbol{\theta}_{\boldsymbol{xy}}^*$ and $\boldsymbol{p}_{\boldsymbol{B}}^{\boldsymbol{C}*}$. Since the initial accelerometer bias is also set to zero, the final estimated accelerometer bias is $\hat{\boldsymbol{b}}_a = \overline{\boldsymbol{b}_a} + \delta\boldsymbol{b}_a^* = \delta\boldsymbol{b}_a^*$. What's more, the gravity in the world frame is adjusted by incorporating perturbation, i.e. $\boldsymbol{g}^{\boldsymbol{W}*} = \boldsymbol{R}_{\boldsymbol{WE}} \cdot Exp(\delta\boldsymbol{\theta}^*)\boldsymbol{g}^{\boldsymbol{E}}$.

4 Camera-IMU Extrinsic Online Estimation

Through the initialization process with camera-IMU extrinsic calibration, an accurate extrinsic parameter could be estimated. However, during the movement of system, the mechanical configuration of the sensor suite changes slightly. Fixed camera-IMU extrinsic parameters can hardly track this change, which leads to system error affecting the tracking accuracy and robustness of the system. With regards to this, we put the camera-IMU extrinsic parameters into the state vectors for online estimation.

4.1 States and Factor Graph Representation

During the motion of the VI-SLAM system, the states to be estimated in each frame include pose, velocity and IMU biases of the sensor suite. On this basis, the camera-IMU extrinsic parameters are also put into the state vector for online estimation. Defining the IMU body frame as the frame to be estimated, the states to be estimated in

each image frame is $\left\{\boldsymbol{p}_{\boldsymbol{B}_i}^{\boldsymbol{W}}, \boldsymbol{R}_{\boldsymbol{WB}_i}, \boldsymbol{v}_{\boldsymbol{B}_i}^{\boldsymbol{W}}, \boldsymbol{b}_{\boldsymbol{g}}^{i}, \boldsymbol{b}_{\boldsymbol{a}}^{i}, \boldsymbol{R}_{\boldsymbol{CB}}, \boldsymbol{p}_{\boldsymbol{B}}^{\boldsymbol{C}}\right\}$. In addition to these, the location of the k th landmark point $\boldsymbol{l}_k^{\boldsymbol{W}} \in \mathbb{R}^3$ is also included in the states. Using factor graph to describe the constraint relationship between these states, the representation of the VI-SLAM system with the camera-IMU extrinsic parameters online estimation is:

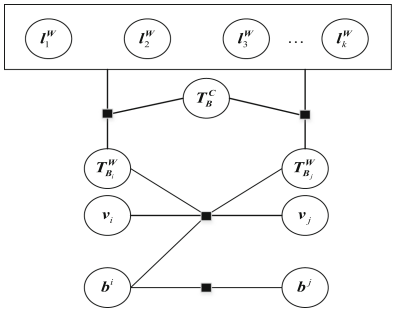

Fig. 3. Factor graph of VI-SLAM system.

As shown in the Fig. 3, the circles represent the variables to be estimated, and squares are the factors. So there are three kinds of constraint in the VI-SLAM system:

(1) Each image pose, camera-IMU extrinsic parameters and landmark point have graph feature position observation constraint;
(2) Poses, velocities and IMU biases of consecutive frames have preintegration constraint of IMU measurements;
(3) IMU biases of consecutive frames have IMU bias random walk constraint.

Therefore, the camera-IMU extrinsic parameters are limited by the graph feature position observation constraint. To estimate it online, it's necessary to construct a nonlinear estimation function of the graph feature position observation constrain.

4.2 Graph Feature Constraint and Its Jacobian

Assuming the states of the frame i is $\left\{\boldsymbol{p}_{\boldsymbol{B}_i}^{\boldsymbol{W}}, \boldsymbol{R}_{\boldsymbol{WB}_i}, \boldsymbol{v}_{\boldsymbol{B}_i}^{\boldsymbol{W}}, \boldsymbol{b}_{\boldsymbol{g}}^{i}, \boldsymbol{b}_{\boldsymbol{a}}^{i}, \boldsymbol{R}_{\boldsymbol{CB}}, \boldsymbol{p}_{\boldsymbol{B}}^{\boldsymbol{C}}\right\}$, and the landmark point k's position in the world frame is $\boldsymbol{l}_k^{\boldsymbol{W}}$. The feature position observed on the frame i of the landmark point k is $\hat{\boldsymbol{p}}_{i,k}$ with the uncertainty of one pixel. Pinhole projection model projects the landmark point $\boldsymbol{l}_k^{\boldsymbol{W}}$ as the pixel $\boldsymbol{p}_{i,k}$.

So the reprojection error $\boldsymbol{e}_{i,k}$ of the graph feature position constraint is:

$$\begin{aligned}\boldsymbol{e}_{i,k} &= \hat{\boldsymbol{p}}_{i,k} - \frac{1}{z_C}\boldsymbol{K}[\boldsymbol{I}_3 \ 0_{3\times 1}]\begin{bmatrix}\boldsymbol{l}_k^C \\ 1\end{bmatrix} \\ &= \hat{\boldsymbol{p}}_{i,k} - \frac{1}{z_C}\boldsymbol{K}[\boldsymbol{I}_3 \ 0_{3\times 1}]\boldsymbol{T}_{CB}\begin{bmatrix}\boldsymbol{R}_{\boldsymbol{WB}_i} & \boldsymbol{p}_{\boldsymbol{B}_i}^{\boldsymbol{W}} \\ 0 & 1\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{l}_k^W \\ 1\end{bmatrix}\end{aligned} \tag{21}$$

$\boldsymbol{K}^{2\times3}$ is the camera intrinsic matrix; the three components of $\boldsymbol{l}_k^C$ (landmark k's position in the camera frame C) is x_C, y_C, z_C; $\boldsymbol{l}_k^C$ can be calculated:

$$\boldsymbol{l}_k^C = \boldsymbol{R}_{CB}\boldsymbol{R}_{WB_i}^{-1}\left(\boldsymbol{l}_k^W - \boldsymbol{p}_{\boldsymbol{B}_i}^{\boldsymbol{W}}\right) + \boldsymbol{p}_{\boldsymbol{B}}^{\boldsymbol{C}} \tag{22}$$

According to Eq. (21), the reprojection error $\boldsymbol{e}_{i,k}$'s Jacobian w.r.t. $\boldsymbol{l}_k^C$ is:

$$\frac{\partial \boldsymbol{e}_{i,k}}{\partial \boldsymbol{l}_k^C} = -\begin{bmatrix} \frac{f_x}{z_C} & 0 & -\frac{f_x x_C}{z_C^2} \\ 0 & \frac{f_y}{z_C} & -\frac{f_y y_C}{z_C^2} \end{bmatrix} \tag{23}$$

f_x, f_y, c_x, c_y are parameters of intrinsic matrix $\boldsymbol{K}$.

In this way, the rejection error $\boldsymbol{e}_{i,k}$ w.r.t. camera-IMU extrinsic translation is:

$$\begin{aligned} \frac{\partial \boldsymbol{e}_{i,k}}{\partial \delta \boldsymbol{p}} &= \frac{\partial \boldsymbol{e}_{i,k}}{\partial \boldsymbol{l}_k^C} \cdot \frac{\partial \boldsymbol{l}_k^C}{\partial \delta \boldsymbol{p}} \\ &= -\begin{bmatrix} \frac{f_x}{z_C} & 0 & -\frac{f_x x_C}{z_C^2} \\ 0 & \frac{f_y}{z_C} & -\frac{f_y y_C}{z_C^2} \end{bmatrix} \boldsymbol{R}_{CB} \end{aligned} \tag{24}$$

Similarly, rejection error $\boldsymbol{e}_{i,k}$ w.r.t. camera-IMU extrinsic rotation is:

$$\begin{aligned} \frac{\partial \boldsymbol{e}_{i,k}}{\partial \delta \boldsymbol{\theta}} &= \frac{\partial \boldsymbol{e}_{i,k}}{\partial \boldsymbol{l}_k^C} \cdot \frac{\partial \boldsymbol{l}_k^C}{\partial \delta \boldsymbol{\theta}} \\ &= \begin{bmatrix} \frac{f_x}{z_C} & 0 & -\frac{f_x x_C}{z_C^2} \\ 0 & \frac{f_y}{z_C} & -\frac{f_y y_C}{z_C^2} \end{bmatrix} \left[\boldsymbol{R}_{CB}\boldsymbol{R}_{WB_i}^{-1}\left(\boldsymbol{l}_k^W - \boldsymbol{p}_{\boldsymbol{B}_i}^{\boldsymbol{W}}\right)\right]^{\wedge} \boldsymbol{R}_{CB} \end{aligned} \tag{25}$$

At this point, the camera-IMU extrinsic parameters can be estimated online based on the residual constraint and the Jacobian.

5 Experimental Evaluation

In this section, performances of our VI-SLAM algorithm with camera-IMU extrinsic calibration and online estimation are estimated on the EuRoC dataset which provides accurate position ground-truth and camera-IMU extrinsic parameters. Eleven sequence, recorded with a Micro Aerial Vehicle (MAV), are divided into three levels: simple, medium and difficult according to different speeds of the aircrafts, illumination, image blur and environment texture. All the experiments are carried out with an Intel CPU i7-5500U (3.0 GHz) laptop computer with 4 GB RAM.

5.1 Implementation Details

The extrinsic rotation calibration is placed in the Tracking thread of the ORB-SLAM, because it's easily excited. The rest the initialization method is implemented in the Local Mapping thread, between the Local BA module and the Local Keyframes Culling module. Considering there are not enough observations to limit the camera-IMU extrinsic parameters in the tracking thread, the camera-IMU extrinsic online estimation is executed in the Local BA module. Since the convergence judgment condition is set only for the extrinsic rotation calibration, the time of the remaining initialization method is set to 23 s.

5.2 Initialization Results

Under the experimental condition of this paper, the initialization and camera-IMU extrinsic estimation results are evaluated using the V2_01_easy sequence of the EuRoC dataset. Figures 4 and 5 show the process of camera-IMU extrinsic rotation calibration. The singular values to Eq. (8) become larger as time goes by (Fig. 4). When the second smallest singular value σ_2 reaches the set threshold $\sigma_{thr} = 0.25$, the process is achieved. A convergence plot of the yaw, pitch and roll can be found in Fig. 5 to their benchmark $[89.147953°,\ 1.476930°,\ 0.215286°]$.

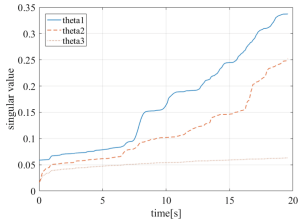

Fig. 4. Time varied singular value.

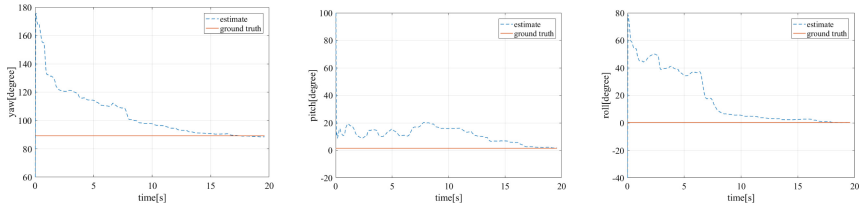

Fig. 5. Process of extrinsic rotation calibration.

The process of camera-IMU extrinsic translation, gyroscope bias, accelerometer bias, scale factor and gravity in world frame is shown in Figs. 6, 7 and 8. Each state quantity begins to converge between 5 s and 10 s after running. The extrinsic translation calibrated will have a few centimeters of error in each axis to its standard $[-0.021, -0.064, 0.009]$m.

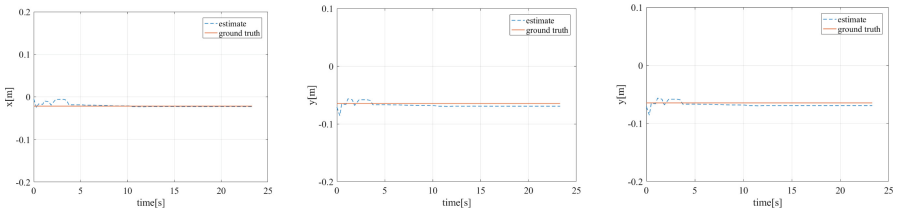

Fig. 6. Process of extrinsic translation calibration.

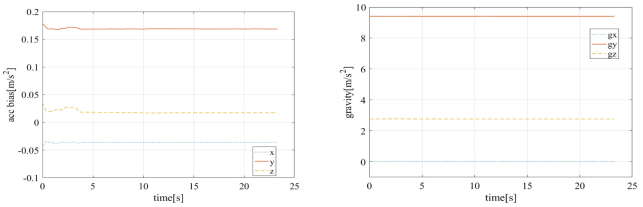

Fig. 7. Process of gyroscope bias and accelerometer bias calibration.

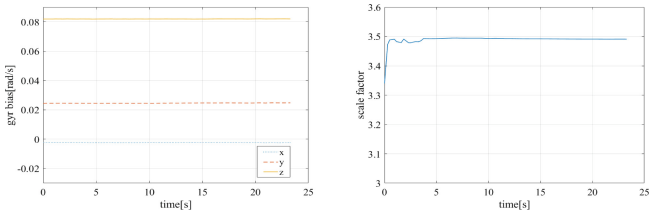

Fig. 8. Process of gravity and scale factor calibration.

5.3 Extrinsic Estimation Results

Figures 9 and 10 illustrate the online estimation process of the camera-IMU extrinsic parameters. During the online estimation process, the deviation of the extrinsic rotation in three axial directions fluctuates within 0.5° and the deviation of the extrinsic translation in three axial directions fluctuates within 0.02 m.

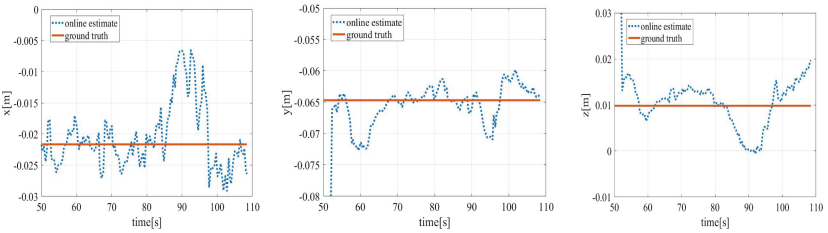

Fig. 9. Process of extrinsic rotation online estimation.

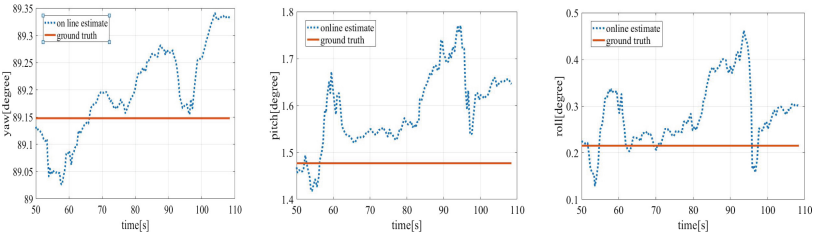

Fig. 10. Process of extrinsic rotation calibration.

Using the V2_01_easy dataset, ten trials for our method, linear and nonlinear processes of VINS-Mono are conducted in Fig. 11. It is observed that the extrinsic rotation calibrated from our method and VINS-Mono' linear process deviates from the benchmark greatly, because we both do not take the gyroscope bias into account. However, our calibrated extrinsic translation is more consistent and accurate than VINS-Mono's linear estimation as we considering the influence of IMU bias. After our online estimation of the extrinsic parameters and VINS-Mono's nonlinear optimization, both camera-IMU extrinsic rotation and translation achieve high precision and consistency. The estimated error of the camera-IMU extrinsic rotation and translation obtained by the method proposed by use, is within 0.5° and 0.02 m separately.

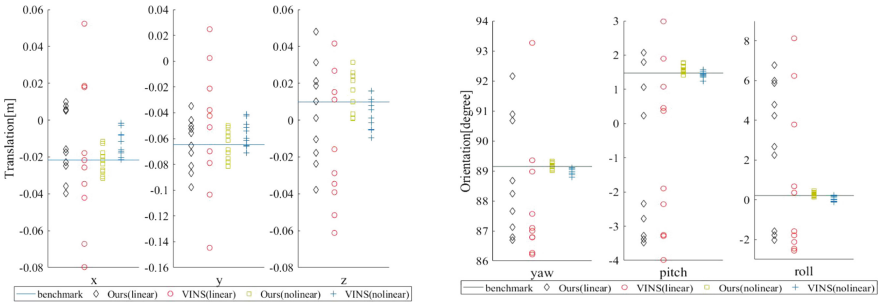

Fig. 11. Process of extrinsic rotation calibration.

6 Conclusion

In this paper, we propose a VI-SLAM algorithm with camera-IMU extrinsic automatic calibration and online estimation without knowing the mechanical configuration of the sensor suite. Compare to VIORB which need prior precise extrinsic calibration, our method is a plug-and-play solution for mobile robots. Through the experiment of EuRoC dataset, the performance of the proposed algorithm in camera-IMU extrinsic calibration and online estimation is verified. The error of the estimated extrinsic rotation and translation is within 0.5° and 0.02 m separately. Compare to VINS-Mono, our method achieves comparable or higher precision and consistency. A limitation of our method is the long time (about 35 s) for the convergence of the extrinsic calibration. To overcome this, we plan to set judgment of convergence in our future work.

References

1. Martinelli, A.: Vision and IMU data fusion: closed-form solutions for attitude, speed, absolute scale and bias determination. IEEE Trans. Robot. **28**(1), 44–60 (2012)
2. Liu, H., Wang, Z., Chen, P.: Feature points selection with flocks of features constraint for visual simultaneous localization and mapping. Int. J. Adv. Robot. Syst. **14**(1), 1–11 (2016)
3. Lin, Y., Gao, F., Qin, T., et al.: Autonomous aerial navigation using monocular visual-inertial fusion. J. Field Robot. **35**(1), 23–51 (2017)
4. Li, M., Mourikis, A.: Improving the accuracy of EKF-based visual-inertial odometry. In: 2012 IEEE International Conference on Robotics and Automation, pp. 828–835. IEEE (2012)
5. Tanskanen, P., Naegeli, T., Pollefeys, M., et al.: Semi-direct EKF-based monocular visual-inertial odometry. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 6073–6075. IEEE (2015)
6. Leutenegger, S., Lynen, S., Bosse, M., et al.: Keyframe-based visual-inertial odometry using nonlinear optimization. Int. J. Robot. Res. **34**(3), 314–334 (2015)
7. Lin, H., Lv, Q., Wang, G., et al.: Robust stereo visual-inertial SLAM using nonlinear optimization. Robot **40**(6), 911–920 (2018)
8. Rehder, J., Siegwart, R.: Camera/IMU calibration revisited. IEEE Sens. J. **17**(11), 3257–3268 (2017)
9. Li, M., Mourikis, A.: High-precision, consistent EKF-based visual-inertial odometry. Int. J. Robot. Res. **32**(6), 690–711 (2013)
10. Dong-Si, T.C., Mourikis, A.I.: Estimator initialization in vision-aided inertial navigation with unknown camera-IMU calibration. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1064–1071. IEEE, Algarve (2012)
11. Yang, S.: Monocular visual-inertial state estimation with online initialization and camera-IMU extrinsic calibration. IEEE Trans. Autom. Sci. Eng. **14**(1), 39–51 (2017)
12. Lupton, S.: Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions. IEEE Trans. Robot. **28**(1), 61–76 (2012)
13. Martinelli, A.: Closed-form solution of visual-inertial structure from motion. Int. J. Comput. Vis. **106**(2), 138–152 (2014)
14. Qin, T., Li, P., Shen, S.: VINS-mono: a robust and versatile monocular visual-inertial state estimator. IEEE Trans. Robot. **34**, 1–17 (2018)

15. Huang, W., Liu, H.: Online initialization and automatic camera-IMU extrinsic calibration for monocular visual-inertial SLAM. In: 2018 IEEE International Conference on Robotics and Automation, pp. 5182–5189. IEEE, Brisbane (2018)
16. Mur-Artal, T.: Visual-inertial monocular SLAM with map reuse. IEEE Robot. Autom. Lett. **2**(2), 796–803 (2016)
17. Forster, C., Carlone, L., Dellaert, F., et al.: On-manifold preintegration for real-time visual-inertial odometry. IEEE Trans. Robot. **33**(1), 1–21 (2015)