Measurement Science and Technology

PAPER

FDO-Calibr: visual-aided IMU calibration based on frequency-domain optimization

To cite this article: Yunfei Liu *et al* 2023 *Meas. Sci. Technol.* **34** 045108

IOP Publishing

Measurement Science and Technology

Meas. Sci. Technol. **34** (2023) 045108 (10pp)

https://doi.org/10.1088/1361-6501/acadfb

FDO-Calibr: visual-aided IMU calibration based on frequency-domain optimization

Yunfei Liu[1,2], Zhitian Li[1], Linhui Xiao[3], Shuaikang Zheng[1,2], Pengcheng Cai[1,2], Haifeng Zhang[1,2], Pengcheng Zheng[1,2] and Xudong Zou[1,2,4,*]

[1] The State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, People's Republic of China

[2] School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, People's Republic of China

[3] National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, People's Republic of China

[4] QiLu Research Institute, Aerospace Information Research Institute Chinese Academy of Sciences, Jinan, People's Republic of China

E-mail: zouxd@aircas.ac.cn

Received 23 September 2022, revised 19 December 2022
Accepted for publication 22 December 2022
Published 23 January 2023

Abstract

Nowadays, multi-sensor fusion technology is a fundamental prerequisite to achieve highly autonomous robots and robustness. Many studies have been conducted, such as visual–inertial odometry (VIO), integrated navigation, and LiDAR–inertial odometry. Typically, for VIO, gratifying results have been achieved, ascribing to the complementary sensing capabilities of inertial measurement units (IMUs) and cameras. However, this work mainly focuses on the fusion of visual and inertial data, while the IMU error is less considered, especially for low-cost or poorly calibrated microelectromechanical system (MEMS) IMU. Such errors may have a significant effect on the VIO performance. In this study, we compensated for the IMU using camera assistance. The key characteristic of the method is that we optimize the compensation parameters (scale factor) from coarse to fine by combining the time domain with the frequency domain. The proposed method is to use the time-domain and frequency-domain optimization to suppress large noise in the dynamic calibration process of the extremely low-cost sensor platform. The effectiveness of this method is validated through experiments and simulations. The minimal calibration error (0.46%) is commensurate with the advanced work. By feeding the compensated IMU into the VIO algorithm, the localization accuracy is improved by 9% to 15%. This method improves the performance in the VIO algorithm, which is equipped with the low-cost or poorly calibrated MEMS IMU and reduces the hardware and deployment costs of the system.

Keywords: low-cost MEMS IMU, sensor calibration, frequency optimization

(Some figures may appear in colour only in the online journal)

[*] Author to whom any correspondence should be addressed.

1. Introduction

In recent years, with the rapid development of autonomous driving [1], augmented reality [2], virtual reality [3], etc, how to obtain the pose estimation of ego-motion accurately is a crucial prerequisite. To meet these emerging application requirements, many positioning methods have been developed, such as Global Positioning System (GPS), inertial navigation [4], and vision-based positioning. In addition, there are emerging methods of multimodal sensor localization, such as floor plan [5], Wi-Fi [5, 6], Bluetooth [7], ultrawide band [8], visible light communication [9], etc. Facing these various positioning technologies, inertial navigation is a mainstream and reliable positioning method, which plays an important role in GPS-denied environments, such as indoor, high-rise buildings, and exoplanet detection. With the emergence of microelectromechanical system (MEMS) technology in recent years, the MEMS-based inertial measurement unit (IMU) has become an important branch [10]. They are nearly everywhere from life-saving airbag accelerometers to missile guidance. MEMS-based IMU, which is widely used in navigation, has been gradually becoming the key sensors in the field of autonomous driving, such as positioning, motion compensation, etc. MEMS IMU with the outstanding advantages of high output frequency (100 Hz to 1000 Hz), low power consumption, and small size is a micro-unit composed of a triaxial accelerometer and a triaxial gyroscope [10]. It can sense acceleration and angular velocity. By integrating the measured acceleration and angular velocity, the position and attitude information of the platform can be obtained. Compared with the camera, IMU is an interoceptive sensor that is not susceptible to the environment and can locate actively. These characteristics make inertial navigation system (INS) an essential positioning technology for integrated navigation and sensor fusion.

Unfortunately, owing to the imperfect manufacturing process and physical characteristics of the sensor, the true measurements of MEMS IMU are contaminated by system errors, random noise, packaging process, interface circuit noise, temperature changes, etc. Noise, bias, scale factor, and installation errors are commonly used to characterize these inaccuracy terms. At the same time, these error terms may change due to mechanical shocks, temperature, and other factors. For IMUs used in professional fields, such as aviation, navigation, and missiles, fine calibration or compensation is completed by the factory, and these IMUs have high-quality performance. They are often not suitable for widespread use in consumer-grade drones, toys, robot platforms with large-scale deployment, and other scenarios on account of their high cost. Cheaper MEMS IMUs are usually used to achieve positioning and navigation, but they are often poorly calibrated. The existence of these errors often makes the cumulative error of pose estimation become larger and larger, resulting in lower confidence. Therefore, the calibration of these deterministic error terms is a crucial part for the long-term stability of localization [11].

On the other hand, due to the richness of visual information, related technologies are also continuous expansion in many fields. Specifically, vision-based positioning methods called visual odometry (VO) have been gradually developed. In VO, the pose of the camera can be directly estimated by the information from adjacent images and the method of multi-view geometry when the robots or platforms are in an unknown environment [12]. Compared with the IMU, the camera that senses and relies on the environment is an exteroceptive sensor and can locate passively. It has the advantage of small drift, and the long-term stability is better than the positioning algorithm based on MEMS IMU. Nevertheless, it is sensitive to fast motion, illumination changes, and occlusion. So, the fusion localization by leveraging the camera and IMU has become a development trend. Recently, a growing trend of visual–inertial odometry (VIO)-related algorithms [13–20] has emerged, and significant progress has been made in robustness and accuracy [21]. However, the current main research still focuses on the fusion of visual and inertial information, and less consideration is given to errors of the IMU.

To this end, we have conducted multi-scenario experiments based on the different grades of IMUs in a previous work [22] and found that the low-cost or poorly calibrated sensors (MPU6050 [23]) would make the VIO's positioning accuracy not as superior as that of high-performance IMUs and sometimes even worse than visual-only positioning. Based on the complementary advantages of VO and MEMS IMU, we propose a vision-aided IMU compensation algorithm for low-cost IMUs (MPU6050).

From the perspective of signal processing, when the IMU is in motion state, the output value of the IMU is composed of the motion sensing signal and the noise signal. We have identified these two parts through Fourier transformation and removed the influence of noise in the frequency optimization process. Based on this, we named the method 'FDO-Calibr'. We use this method to compensate the scale factor of the MEMS IMU whose noise is a significant error source. The compensated IMU improves the positioning accuracy and has potential to reduce the hardware and deployment costs of the system. The main contributions of this study are summarized as follows.

The method introduced in this paper is to use the time-domain and frequency-domain optimization to suppress the large noise in the dynamic calibration process of the extremely low-cost sensor platform.

The rest of this paper is structured as follows. In section 2, related work about IMU calibration or compensation is discussed. In section 3, the framework flow of the algorithm, which is divided into six parts, is presented. In section 4, the method of the compensation algorithm is analyzed and discussed. Relevant simulation and experiment are discussed in section 5. Finally, the paper is concluded in section 6.

2. Related work

Several authors [24–26] utilized gravity signal as a stable reference when an accelerometer is at static periods. The key lies in how to better linearize the constraints, or there will be convergence problems. However, their method is only suitable for calibrating the MEMS accelerometers. The calibration of the gyroscope is achieved by means of a three-axis turntable, which can provide high-precision angular-velocity information or angle information. However, the turntable is expensive, complicated to operate, and bulky enough, which make it impractical and difficult to use.

The calibration method proposed by Qureshi and Golnaraghi [27] and Tedaldi *et al* [28] can calibrate the IMU without external equipment. For an accelerometer, the algorithm detected the time period clusters, in which the accelerometer is in a static state under different attitudes, and then the cost function was established with a constant gravity norm value of g. When the internal parameters of the accelerometer were solved, the angle calculated by the accelerometer and the angle integrated by the gyroscope were used to establish a cost function about the gyroscope. The calibration of the gyroscope was coupled with accelerometer calibration. The calibration accuracy of the gyro triad strongly depends on that of the accelerometer triad. Li *et al* [29] used pseudo-observations to replace the GPS measurements in the loosely coupled GPS/INS integrated systems based on the assumption that 'if an IMU were rotating strictly around its measurement center, its position would remain constant, and its linear velocity would be zero'. In addition, Xiao *et al* [30] employed the time-domain optimization method to iteratively solve the IMU parameters and positioning-related states together. The state vector to maintain was too large, and there is no advantage of light weight. The Kalibr algorithm developed by Rehder *et al* [31] realized the joint calibration of the camera and the IMU. However, it needs the help of a calibration board placed in the environment, which is cumbersome to operate. Syed *et al* [32] presented a multi-position calibration method to calibrate the IMU but needed a high-precision turntable. In this research work, the authors paid little attention to the noise when the sensors were in dynamic motion process. Noise is a significant error source for the MEMS IMU sensors, especially for low-cost MEMS IMUs.

Based on these, we used the extremely low-cost MEMS IMU (MPU6050), which is less than \$0.5, in our experiment to verify the effectiveness of the method. Our method can suppress the noise that occurs when the sensor is dynamically calibrated. In addition, from the perspective of application, the experiment that the compensated IMU improves the positioning accuracy of multi-sensor fusion is also verified.

3. System overview

The specific framework of the system is shown in figure 1. To complete the error compensation of low-cost MEMS IMU, six steps need to be executed. The entry of the pipeline is to get the IMU data stream first. Then, the image stream is fed into the VO algorithm to obtain the pose information of the camera. The motion information based on vision is obtained by interpolation and Kalman Smoothing (KS), which is the third step. The motion information includes the linear acceleration and angular velocity of the platform. The fourth step is to spatially and temporally align the visual and inertial motion data and remove the bias. In the fifth step, the cost function is constructed to perform the optimization solution in the time domain first. Due to the existence of nonlinear factors in the frequency-domain optimization, the time-domain optimization result of this step is used as the initial value of the frequency-domain optimization in the fifth step. This step completes the coarse-to-fine optimization. In the last step, the compensated IMU data are fed into the VIO algorithm to verify the effect of the visual-assisted IMU compensation algorithm on the VIO positioning accuracy.

To better express the association among the data of different frames, we define the relevant notations and frame transformation, in which w represents the world frame, c represents the camera frame, and b represents the IMU frame body. For example, $(\boldsymbol{X})_{\mathrm{b}}$ represents the variable vector $\boldsymbol{X}$ in the body frame, and $\boldsymbol{T}_{\mathrm{wb}}$ represents the Euclidean transformation from the body frame to the world frame. V represents the information obtained by visual methods, I represents the information obtained by inertial MEMS IMU methods, and V, I is used to distinguish the different information sources. $\boldsymbol{X}^{V}_{\mathrm{c/b}}$ represents the visual vector information $\boldsymbol{X}$ in the camera or body frame. $X^{I}_{\mathrm{c/b}}$ represents the inertial vector information $\boldsymbol{X}$ in the camera or body frame. We use rotation matrices $\boldsymbol{R}$ or Hamilton quaternions $\boldsymbol{q}$ to represent rotations; $\otimes$ represents the multiplication operator between two quaternions, and $\boldsymbol{q}^{-1}$ represents the inverse operation of a quaternion. Moreover, $\boldsymbol{g}_{\mathrm{w}} = [0, 0, g]^{\mathrm{T}}$ is the gravity vector in the world frame.

4. Calibration methods

4.1. Measurement model of the IMU

The measured value of the IMU will be affected by many error sources, resulting in deviation from the real physical value. The common error items are noise, bias, scale factor, and misalignment. The noise refers to the additive noise caused by various noise sources. The higher the noise is, the larger are the errors caused by IMU measurements. The bias is an error value that slowly shifts between the output value and the input value with the running time. The scale factor represents the ratio between the measured value and the true value. The installation error represents the degree of non-orthogonality among the three axes. The measurement model of the low-cost MEMS IMU is as follows [33],

$$\widehat{a^{I}_{\mathrm{b}}} = T_{\mathrm{a}} a^{I}_{\mathrm{b}} + n_{\mathrm{a}} + b_{\mathrm{a}} \tag{1}$$

$$\widehat{w^{I}_{\mathrm{b}}} = T_{\mathrm{g}} w^{I}_{\mathrm{b}} + n_{\mathrm{g}} + b_{\mathrm{g}}. \tag{2}$$

In these equations, $\widehat{a^{I}_{\mathrm{b}}}, \widehat{w^{I}_{\mathrm{b}}}$ and $a^{I}_{\mathrm{b}}, w^{I}_{\mathrm{b}}$ represent the measurements and true values of acceleration and angular velocity in the body frame, respectively. The accelerometer is used to sense the vector sum of the linear acceleration of motion and the acceleration of gravity. The gyroscope is used to sense the angular velocity of the body; $n_{\mathrm{a}}, b_{\mathrm{a}}$ represents the accelerometer noise and bias, and $n_{\mathrm{g}}, b_{\mathrm{g}}$ represents the gyroscope noise and bias. We assume that the noise in the accelerometer and the gyroscope measurements are Gaussian, that is, $n_{\mathrm{g}} \sim$

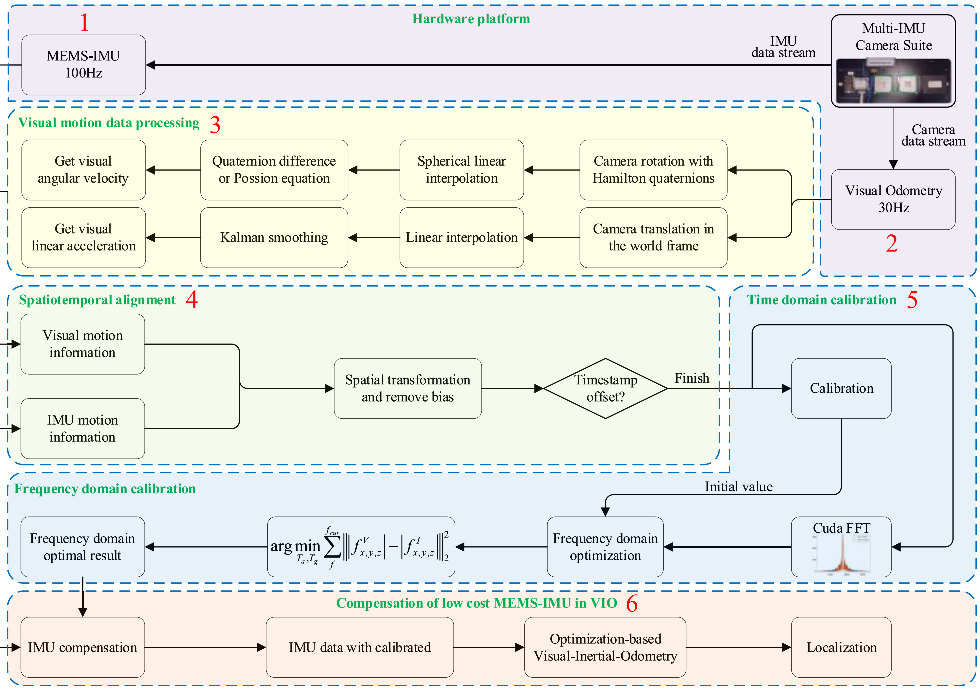

Figure 1. Full pipeline of the proposed IMU calibration with visual-aided IMU compensation algorithm based on frequency-domain optimization.

$N\left(0,\sigma_{\mathrm{g}}^{2}\right)$, $n_{\mathrm{a}} \sim N\left(0,\sigma_{\mathrm{a}}^{2}\right)$. As for the bias, gyroscope bias and acceleration bias are modeled as random walk, whose derivatives are Gaussian, $n_{\mathrm{b_a}} \sim N\left(0,\sigma_{\mathrm{b_a}}^{2}\right)$, $n_{\mathrm{b_g}} \sim N\left(0,\sigma_{\mathrm{b_g}}^{2}\right)$,

$$\dot{b}_{\mathrm{a}} = n_{\mathrm{b_a}}, \dot{b}_{\mathrm{g}} = n_{\mathrm{b_g}}. \tag{3}$$

$T_{\mathrm{a}}, T_{\mathrm{g}}$ represents the product matrix of the scale factor and the cross-axis error of the three-axis accelerometer and gyroscope, which is called compensation matrix or calibration matrix. The main diagonal represents the scale factor, and the other elements are the degree of misalignment among the axes. The specific form is as follows,

$$T_{\mathrm{a}} = \begin{bmatrix} s_x^{\mathrm{a}} & m_{12}^{\mathrm{a}} & m_{13}^{\mathrm{a}} \\ m_{21}^{\mathrm{a}} & s_y^{\mathrm{a}} & m_{23}^{\mathrm{a}} \\ m_{31}^{\mathrm{a}} & m_{32}^{\mathrm{a}} & s_z^{\mathrm{a}} \end{bmatrix}, T_{\mathrm{g}} = \begin{bmatrix} s_x^{\mathrm{g}} & m_{12}^{\mathrm{g}} & m_{13}^{\mathrm{g}} \\ m_{21}^{\mathrm{g}} & s_y^{\mathrm{g}} & m_{23}^{\mathrm{g}} \\ m_{31}^{\mathrm{g}} & m_{32}^{\mathrm{g}} & s_z^{\mathrm{g}} \end{bmatrix}. \tag{4}$$

The goal is to identify the measurement model of low-cost MEMS IMU. Through experiments, it was found that the scale factors of these poorly calibrated MEMS IMUs had a more significant impact on the positioning accuracy of MEMS IMU in multi-sensor fusion. The value of non-diagonal elements was $k \times 10^{-3}$, which was very close to 0 and had less impact on the results. Here, k represents a constant in scientific notation. Therefore, the optimization algorithm of this paper compensates and calibrates the scale-factor error source.

4.2. Visual odometry

The monocular camera that has been widely studied has the advantages of miniaturization, high integration, and low cost. However, the depth information dimension will miss when the monocular camera is projected from the 3D space to the 2D plane of the camera. Therefore, it is often not used in the actual positioning and navigation field. Instead, the Red, Green, Blue and Depth (RGB-D) map and binocular camera can obtain depth information, and the real scale of the overall positioning trajectory can be constructed. In this paper, the VO framework VINS_Fusion [20], which is based on sliding window optimization, is adopted [18]. The binocular camera is introduced in visual-only odometry to obtain the rotation matrix R_{wc} and position vector p_{wc} of the camera. R_{wc} can also be expressed in the form of Hamilton quaternions q_{wc}.

4.3. Visual motion data processing

4.3.1. Visual linear acceleration. In the experiment, the output frequency of the VO is 30 Hz, and the IMU frequency is 100 Hz. Here, the camera output frequency is 60 Hz, but limited by the computing time, the odometry output frequency of the binocular VO is 30 Hz. Due to the low frequency of the camera odometry, the camera position information is linearly interpolated to match the frequency of the IMU. To obtain the visual acceleration, the KS algorithm, which is divided into forward filtering and backward smoothing, is implemented. Because the physical motion law is considered in the prediction model and the observation value of the camera position information is used in the measurement model, better acceleration information can be obtained through the state estimation algorithm. The state vector of the KS at time step k is given by,

$$X_k = \left[\begin{array}{ccc} p_k & v_k & a_k \end{array}\right]. \tag{5}$$

In equation (5), p_k v_k a_k represent the state vectors for position, velocity, and acceleration, respectively. The prediction process and its covariance estimation are performed first,

$$\begin{aligned}\hat{X}_k &= A_k\tilde{X}_{k-1} + \epsilon_k \\ \hat{P}_k &= A_k P_{k-1} A_k^T + Q_k.\end{aligned} \tag{6}$$

$A_k = \begin{bmatrix} 1 & \Delta t & \frac{1}{2}\Delta t^2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}$ represents the state transition matrix. Δt represents the sampling time, $\widetilde{(\cdot)}$ represents the optimal state estimate, and $\widehat{(\cdot)}$ represents a predicted value. For example, $\tilde{X}_{k-1}$ represents the optimal state at time $k-1$. $\hat{X}_k$ represents the predicted state at time k. Similarly, the matrix P represents the covariance matrix of the optimal state or predicted state. We assume that the noise in the predictive model is Gaussian, $\epsilon_k \sim N(0, Q_k)$. When the measurements of the camera position arrive, the update procedure and the covariance estimation were performed,

$$\begin{aligned}Z_k &= H_k X_k + \delta_k \\ K_k &= \hat{P}_k H_k^{\mathrm{T}}\left(H_k\hat{P}_k H_k^{\mathrm{T}} + R_k\right)^{-1} \\ \tilde{X}_k &= \hat{X}_k + K_k\left(Z_k - H_k\hat{X}_k\right) \\ \tilde{P}_k &= \hat{P}_k - K_k H_k \hat{P}_k.\end{aligned} \tag{7}$$

Z_k represents the observation in the Kalman algorithm, which refers to the camera position p_{wc} in section 4.2. H_k represents the transformation matrix between the state and the measurements. We assume that the noise in measurements is Gaussian, $\delta_k \sim \mathcal{N}(0, R_k)$. K_k represents the optimal weight between the measurements and the predicted state.

Forward filtering completed the forward recursion with time step $[t_0, t_1, \ldots, t_k]$, and backward smoothing completed the backward recursion in $[t_k, t_{k-1}, \ldots, t_0]$. Because all position information of the camera is available at the whole sample time, the influence of noise is reduced by using the KS algorithm to further improve the accuracy of visual acceleration. The formulation of the specific smoothing is as follows,

$$\begin{aligned}\bar{X}_{k+1} &= A_k\tilde{X}_k + \epsilon_k \\ \bar{P}_{k+1} &= A_k\tilde{P}_k A_k^{\mathrm{T}} + Q_k \\ G_k &= \tilde{P}_k A_k^{\mathrm{T}}(\bar{P}_{k+1})^{-1} \\ G_k &= \tilde{P}_k A_k^{\mathrm{T}}(\bar{P}_{k+1})^{-1} \\ P_k^{\mathrm{s}} &= \tilde{P}_k + G_k\left(P_{k+1}^{\mathrm{s}} - \bar{P}_{k+1}\right)G_k^{\mathrm{T}}\end{aligned} \quad . \tag{8}$$

$\bar{X}_{k+1}$ represents the backward predicted state based on $\tilde{X}_k$. $\bar{P}_{k+1}$ represents the covariance matrix of the predicted state. P_k^{s} represents the covariance matrix of the backward smoothing optimal state. G_k represents the optimal weight between the smoothing state and the predicted state. The optimal estimated state X_k^{s} is obtained by the KS algorithm, from which the visual acceleration information $\boldsymbol{a}_{\mathrm{c}}^V$ can be obtained.

4.3.2. Visual angular velocity. The rotation information of the camera can be obtained by VO. Because the frequency of the camera is inconsistent with the IMU, spherical linear interpolation of the quaternion rotation is needed. The visual angular velocity can be solved by the differential of the quaternion. The derivative of the quaternion with respect to time is as follows,

$$\dot{q}_{\mathrm{wc}} = \frac{1}{2}q_{\mathrm{wc}} \otimes \begin{bmatrix} 0 \\ w_{\mathrm{c}}^{\mathrm{V}} \end{bmatrix}. \tag{9}$$

So, there is,

$$\begin{bmatrix} 0 \\ w_{\mathrm{c}}^V \end{bmatrix} = 2q_{\mathrm{wc}}^{-1} \otimes \dot{q}_{\mathrm{wc}}. \tag{10}$$

The element corresponding to the imaginary part is the visual angular-velocity information in the camera frame w_{c}^V.

4.4. Spatial and temporal transformation

Spatiotemporal alignment is an interactive bridge between the IMU data in the body frame and the vision data in the camera frame. For the purpose of compensation, it is necessary to unify the data in different frames into the same frame. Here, we chose the camera frame.

The camera and the IMU module are connected to the same circuit board through a rigid body. In spatial alignment, there is no need to consider translation parameters due to the processing of the angular-velocity and linear-velocity vectors. The rotation matrix R_{cb} can be obtained through the geometric relationship of the frames. For temporal alignment, the angular-velocity information is used to obtain the time offset t_d between vision and inertia based on the golden ratio search algorithm. The cost function about time offset can be constructed,

$$t_d{}^* = \arg\min\sum_t \left\|w_{\mathrm{c}}^V(t) - R_{\mathrm{cb}}w_{\mathrm{b}}^I(t+t_d)\right\|_2^2 \tag{11}$$

where w_{b}^I represents the angular-velocity information of the inertial IMU in the body frame. R_{cb} denotes the transformation from the body frame to the camera frame. The measured values in the body frame can be converted into the camera frame by R_{cb}, as shown in equation (12). Figure 2 shows the spatial and temporal alignment,

$$w_{\mathrm{c}}^I(t) = R_{\mathrm{cb}}w_{\mathrm{b}}^I(t). \tag{12}$$

4.5. Compensating accelerometer and gyroscope

4.5.1. Coarse optimization of the time domain. A. Time-domain optimization of accelerometer compensation matrix: the average value of the stationary period T is taken as the IMU bias $b_{\mathrm{a}}, b_{\mathrm{g}}$, where T is 5 s. Because the acceleration of the inertial device is gravity mixed and the visual linear acceleration is non-observable for the gravity vector $g_{\mathrm{w}} = (0, 0, g)$ in the inertial frame, the inertial acceleration needs to remove the influence of the gravity vector in the inertial frame.

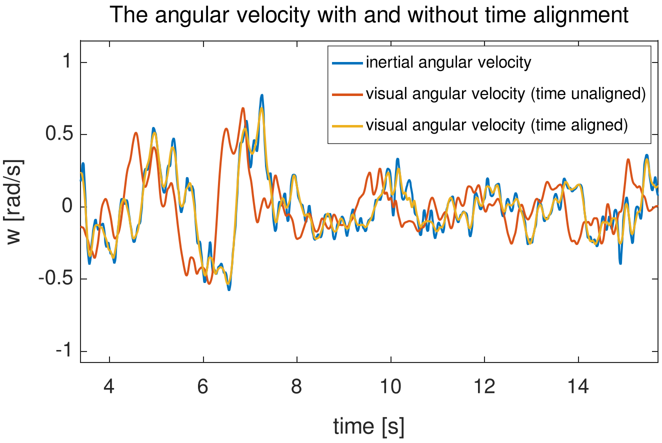

Figure 2. Comparison between with and without temporal alignment. The blue line represents the inertial angular velocity. The red line represents the visual angular velocity without temporal alignment. The yellow line represents the visual angular velocity after temporal alignment.

Otherwise, there will be a large proportion in the frequency-domain optimization, which will affect the optimization results. The time-domain cost function is as follows,

$$\begin{aligned}&\widehat{a_{\mathrm{c}}^{I}}(t)=R_{\mathrm{cb}}\left(\widehat{a_{\mathrm{b}}^{I}}(t)-b_{\mathrm{a}}\right)\\&T_{\mathrm{at}}^{-1*}=\arg\min\sum_{t}\left\|a_{\mathrm{c}}^{V}(t)-T_{\mathrm{at}}^{-1}*\left(\widehat{a_{\mathrm{c}}^{I}}(t)-R_{\mathrm{cw}}(t)\,g_{\mathrm{w}}\right)\right\|_{2}^{2}.\end{aligned}\tag{13}$$

In equation (13), $\widehat{a_{\mathrm{c}}^{I}}$ indicates the inertial acceleration information after bias removal in the camera frame; a_{c}^{V} represents the visual linear acceleration in the camera frame.R_{cw} is the transpose of the camera rotation matrix R_{wc} in section 4.2. R_{cw} converts the gravity in the world frame to the camera frame. In addition, $\|\cdot\|$ represents the norm of a vector. T_{at}^{-1} represents the state to be optimized. The subscript t represents solving T_{a} in equation (1) through the time-domain method. The estimation value of the accelerometer compensation matrix under time-domain optimization can be solved by constructing the optimization factor of this cost function by using Ceres [34].

B. Time-domain optimization of gyroscope compensation matrix: similar to the accelerometers, visual angular velocity is used to assist in compensating for inertial angular-velocity information. The time-domain error cost function is established as follows,

$$\begin{aligned}&\widehat{w_{\mathrm{c}}^{I}}(t)=R_{\mathrm{cb}}\left(\widehat{w_{\mathrm{b}}^{I}}(t)-b_{\mathrm{g}}\right)\\&T_{\mathrm{gt}}^{-1*}=\arg\min\sum_{t}\left\|w_{\mathrm{c}}^{V}(t)-T_{\mathrm{gt}}^{-1}*\widehat{w_{\mathrm{c}}^{I}}(t)\right\|_{2}^{2}.\end{aligned}\tag{14}$$

In equation (14), $\widehat{w_{\mathrm{c}}^{I}}$ represents the inertial angular-velocity information in the camera frame after bias is removed; w_{c}^{V} represents the visual angular-velocity information in the camera frame. T_{gt}^{-1} denotes the optimized state. The estimated value of the gyroscope compensation matrix in the time domain can be calculated by constructing the optimization factor of this cost function by using Ceres.

Through the cost function, the coarse compensation of the IMU is completed under the time domain. The estimated value of the coarse compensation will be used as the initial value of the fine compensation in the frequency domain.

4.5.2. Fine optimization of the frequency domain. In the preceding time-domain optimization, a global optimum is obtained by least-squares optimization. Compared with the time domain, the frequency-domain optimization realizes the separation of low-frequency useful signals and high-frequency noise signals. The optimization accuracy can be further improved by using low-frequency signals. In many measurement models, it is generally assumed that the noise is Gaussian white noise, and the assumption of noise can be ignored by separating in the frequency domain. Based on the method of the frequency-domain optimization proposed by Mustaniemi *et al* [35], we further applied it to the compensation of low-cost or poorly calibrated MEMS IMU and completed the fine optimization in the frequency domain based on vision and inertia. Unlike the IMU-preintegration method [36], the frequency-domain information can be made full use of by obtaining the visual and inertial motion information.

A. Fine optimization for accelerometer in frequency domain: because the frequency-domain transformation has nonlinear effects, the time-domain optimization solution in section 4.5.1 needs to be used as the initial value of the frequency-domain optimization. For the accelerometer, the cost function constructed in the frequency domain is as follows,

$$\begin{aligned}&F_{ai}^{V}=\mathcal{F}\left(a_{\mathrm{c}}^{V}(t)\right),\ i=x,y,z\\&F_{ai}^{I}=\mathcal{F}\left(T_{\mathrm{af}}^{-1}*\left(\widehat{a_{\mathrm{c}}^{I}}(t)-R_{\mathrm{cw}}(t)\,g_{\mathrm{w}}\right)\right),\ i=x,y,z\\&T_{\mathrm{af}}^{-1*}=\arg\min\sum_{f}^{f_{\mathrm{cut}}}\left\|\left|F_{ai}^{V}\right|-\left|F_{ai}^{I}\right|\right\|_{2}^{2},\ i=x,y,z\end{aligned}\tag{15}$$

where F_{ai}^{V} represents the complex value of the visual acceleration of each axis $a_{\mathrm{c}}^{V}(t)$ after Fourier transform $\mathcal{F}(\cdot)$. F_{ai}^{I} represents the complex value of the inertial acceleration of each axis after the Fourier transform. In addition, $|\cdot|$ represents the norm value of the Fourier transform result. When the amplitude of Fast Fourier transform (FFT) of the IMU is obviously smaller and in a flat band, these signals will be considered as the noise. In the equation, f_{cut} indicates the selected frequency in frequency-domain optimization, which was selected as 10.2 Hz in the paper. T_{af}^{-1} represents the state to be optimized in the frequency domain. The subscript f represents further using the frequency-domain method to solve T_{a} in equation (1) based on the time-domain results. Constructing the frequency-domain optimization factor of this cost function by using Ceres can solve the optimization value under the frequency domain.

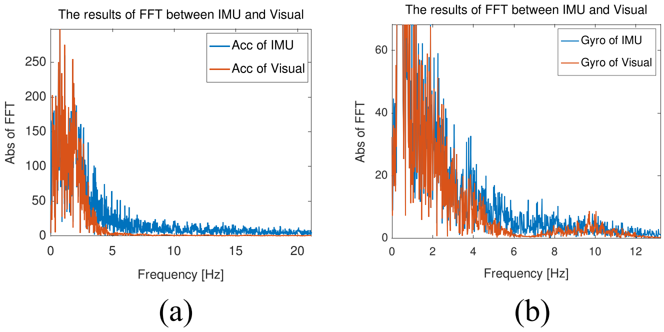

Figure 3. Results of FFT between IMU and visual: (a) visual and inertial acceleration information in the frequency domain. (b) Visual and inertial angular-velocity information in the frequency domain. Where Acc represents the accelerometer, Gyro represents the gyroscope, and Abs represents the norm of the FFT result.

B. Fine optimization for gyroscope in frequency domain: for the gyroscope, the cost function constructed in the frequency domain is as follows,

$$
\begin{aligned}
F_{gi}^{V} &= \mathcal{F}\left(w_{\mathrm{c}}^{V}(t)\right),\ i=x,y,z \\
F_{gi}^{I} &= \mathcal{F}\left(T_{\mathrm{gf}}^{-1} * \widehat{w_{\mathrm{c}}^{I}}(t)\right),\ i=x,y,z \\
T_{\mathrm{gf}}^{-1} &= \arg\min \sum_{f}^{f_{\mathrm{cut}}} \left\| \left|F_{gi}^{V}\right| - \left|F_{gi}^{I}\right| \right\|_2^2,\ i=x,y,z
\end{aligned} \tag{16}
$$

where F_{gi}^{V} represents the complex value of the visual angular velocity of each axis $w_{\mathrm{c}}^{V}(t)$ after Fourier transform. F_{gi}^{I} represents the complex value of the inertial angular velocity of each axis after the Fourier transform. T_{gf}^{-1} represents the state to be optimized in the frequency domain.

Figure 3 shows the distribution of visual and inertial information in the frequency domain. In the whole frequency band, the motion information is mainly concentrated in the low-frequency band, while the noise signal occupies a high-frequency band.

To quickly process the Fourier transform in the Ceres optimization process, the cuFFT module based on Compute Unified Device Architecture (CUDA) platform is introduced to speed up the optimization solution process, which is 10^4 times faster than the traditional FFT. The cuFFT module is a fast Fourier transform algorithm running on the cuda platform.

5. Simulation and experiment

5.1. Simulation

To verify the effectiveness of the proposed algorithm, simulation tests have been performed. We generated the measured value information of the IMU and camera odometry based on the measurement model in section 4.1 to try to match the actual data. A piece of data about 45 s was generated, and the true value was set to (1.04 1.05 1.06) in diagonal matrix. The data in the time and frequency domains are shown in figure 4.

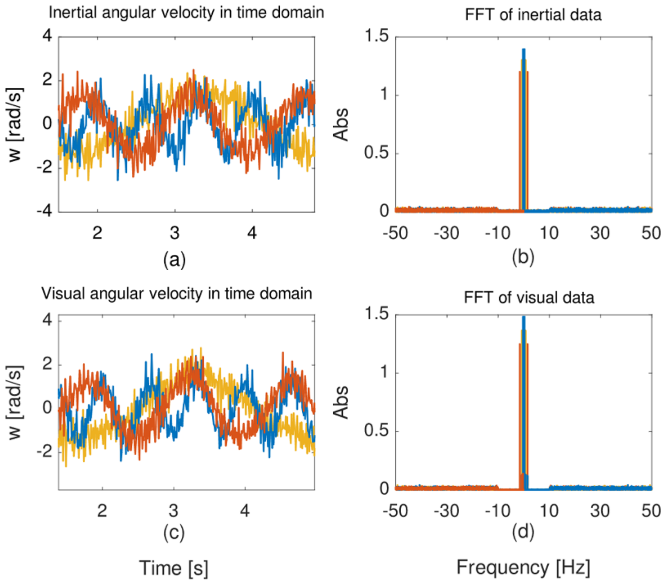

Figure 4. Simulation data for the visual and inertial angular velocities: (a) and (c) represent the time-domain data polluted by noise, whereas (b) and (d) represent the frequency-domain segmentation result.

Table 1. Simulations: summary of calibration parameter results and RE. RE represents reference value. The abbreviation 'TO' represents the time-domain optimization. The abbreviation 'FO' represents fine optimization in the frequency domain. The third column represents the results of TO. The fourth column represents the RE of TO. The fifth column represents the results of FO. The sixth column represents the RE of TO.

Simulation	RE	TO		FO	
s_x^{a}	1.04	0.767	26.20%	1.024	1.56%
s_y^{a}	1.05	0.891	15.16%	1.033	1.66%
s_z^{a}	1.06	1.022	3.55%	1.057	0.24%
s_x^{g}	1.04	0.973	6.44%	1.033	0.64%
s_y^{g}	1.05	1.004	4.37%	1.036	1.31%
s_z^{g}	1.06	1.031	2.76%	1.050	0.90%

Relative error (RE) represents the error relative to the true value, $\tilde{x}$ represents the estimated value, and x represents the ground truth value for reference,

$$\mathrm{RE} = \frac{|\tilde{x} - x|}{x} \times 100\%. \tag{17}$$

Table 1 shows the calibration results about the three main calibration parameters of the low-cost MEMS IMU. It can be found that the compensation parameters are reduced from the time-domain error (26.20%, 15.16%, and 3.55%) to the frequency-domain error (1.56%, 1.66%, and 0.24%) for the accelerometer, and the compensation parameters are reduced from the time-domain error (6.44%, 4.37%, and 2.76%) to the frequency-domain error (0.64%, 1.31%, and 0.90%) for the gyroscope. Through the simulation, the frequency-domain compensation algorithm proposed in this paper reduces the impact of high-frequency noise on optimization and improves

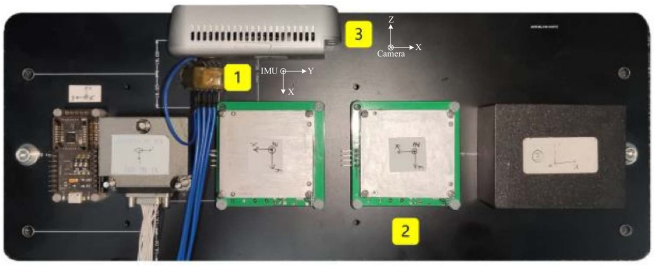

Figure 5. Sensor suite of multi-IMU camera. It contains two forward-looking global shutter cameras (RealSenseD435i) with 640 × 480 resolution and different grades of IMUs. No. 1, No. 2 and No. 3 represent MPU6050, ADIS16490 and RealSenseD435i respectively.

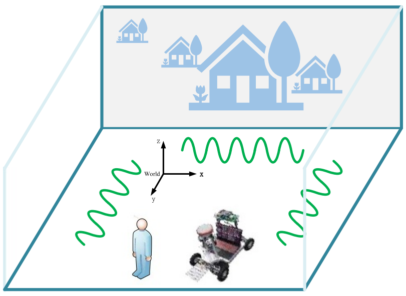

Figure 6. Calibrate settings to obtain excitation for different poses by handheld or remote-controlled vehicles. The wavy line represents the excitation of the irregular motion to obtain the measurements.

the estimation accuracy of the error parameters we are concerned about.

5.2. Experimental verification

To further verify the effectiveness of the algorithm, a multi-IMU camera platform was built to perform the experiments. The sensor suite we used is shown in figure 5. The MEMS IMU named MPU6050 represented the IMU module that needed to be compensated in the experiment. The binocular camera, which was well corrected by Kalibr [37], was used as the input of the VO to obtain the pose information of the camera. The frequencies of the IMU and the camera were 100 Hz and 60 Hz, respectively. The image frequency should be as high as possible so that the interpolation process will not cause too much information missing due to the low frequency.

To be compensated for each axis of the MEMS IMU, rotation and variable-speed translation around the three axes are performed to obtain the measurement excitation. In our experiments, we conducted these operations with a handheld sensor suite in a texture-rich scene. Because it is easy to realize variable-speed movement by handheld, it can meet the requirements of measurement excitation. In addition to handheld devices, we can also complete data collection by remote-controlled vehicles walking different routes through the schematic diagram shown in figure 6.

Table 2. Summary of calibration parameter results and RE for MPU6050. The abbreviation explanation of the first row is consistent with that in table 1.

MPU6050	RE	TO		FO	
s_x^a	0.975	1.021	4.75%	0.964	1.14%
s_y^a	0.905	1.030	13.8%	0.911	0.67%
s_z^a	1.024	1.010	1.42%	0.984	3.89%
s_x^g	0.965	1.022	5.97%	0.969	0.46%
s_y^g	0.989	0.993	0.47%	1.005	1.61%
s_z^g	0.974	1.006	3.29%	0.991	1.82%

By executing our algorithm of coarse optimization in time domain and fine optimization in frequency domain, the calibration of MPU6050 was completed. Because the ADIS16490 [38] has a very high precision in MEMS IMU, the results compensated by ADIS16490 were used for reference to verify the validity of the experimental results. The specific results are summarized in table 2. The average calibration error of the scale factor is 1.6%. From table 2, it can be found that after the frequency-domain optimization, the overall error is smaller. Compared with some advanced work, the minimum calibration error of 0.46% in our work is commensurate with the 0.34% in [27] and 400– 700 ppm in [29]. Our method can suppress the noise that occurs when the sensor is dynamically calibrated.

5.3. Indoor fusion positioning experiment

Based on the VINS_Fusion [20] algorithm framework, the compensated MPU6050 was used for VIO fusion positioning to verify the improvement of the localization accuracy. As for the test experiment, loop-closure detection was not enabled because the detection was a stronger constraint for the compensated MPU6050. Enabling the detection would reduce the constraints of the IMU on the fusion framework.

To evaluate the accuracy, the VINS_Mono algorithm framework [18] was used as a ground truth because of the higher positioning accuracy in monocular and IMU [20]. ADIS16490, which is expensive and has the highest nominal accuracy, was used for the IMU sensor. The left camera of RealSenseD435i [39] was used for monocular image. As for the reference experiment, the loop-closure detection was enabled for globally consistent trajectories. These operations guaranteed that the reference value had the highest accuracy about positioning.

We performed seven experiments through the handheld device in the indoor environment of the office. Figure 7 shows the indoor environment of the office. The comparison of trajectory of one of the sequences named room6 is shown in figure 8.

To further characterize the situation of the positioning error, the box plot of the error is shown in figure 9. It can be demonstrated that the compensated IMU reduces the error in multi-sensor fusion.

Figure 7. Images recorded during the indoor experiment.

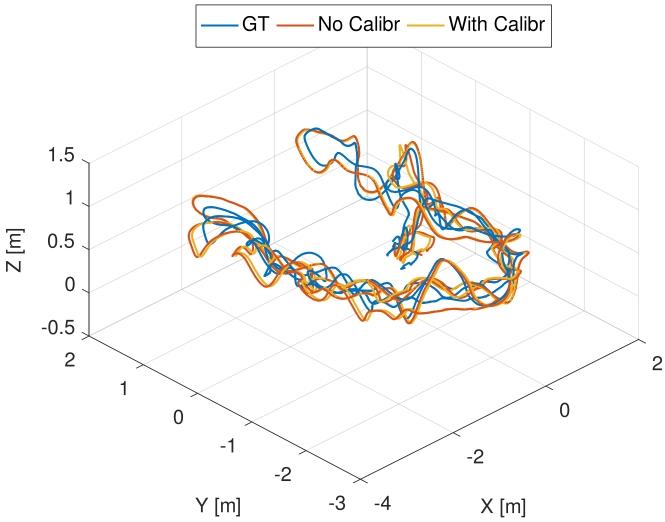

Figure 8. Comparison of trajectory of room6. The blue line represents the ground truth (GT) value. The red line represents the VIO trajectory without the compensated IMU. The yellow line represents the VIO trajectory of the IMU after compensation.

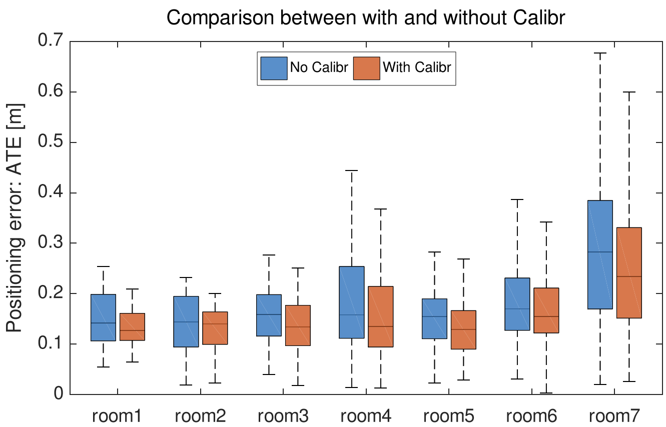

Figure 9. Box plot for positioning error. ATE represents the absolute trajectory error.

Moreover, the Root Mean Square Error (RMSE) is summarized in table 3, and the accuracy is improved by $9\% - 15\%$. The result verifies the effectiveness of the compensation IMU algorithm and the improvement of the multi-sensor fusion positioning accuracy.

Table 3. RMSE of the indoor fusion positioning based on the compensated MPU6050 and camera. 'Time' represents the duration of the experiment. 'Traj.L' represents the distances traveled. 'No Calibr' represents the RMSE results of multi-sensor fusion positioning using an uncompensated IMU, and 'with Calibr' represents the RMSE results of multi-sensor fusion positioning using the compensated IMU. 'Improvements' represents the improvement of positioning accuracy.

Indoor	Time (s)	Traj.L (m)	No Calibr (m)	With Calibr (m)	Improvements (%)
Room1	61.8	22.6	0.1585	0.1396	11.9
Room2	91.7	26.0	0.1519	0.1376	9.4
Room3	56.1	20.1	0.1690	0.1469	13.1
Room4	88.5	29.8	0.2076	0.1761	15.2
Room5	102.5	31.3	0.1621	0.1412	12.9
Room6	153.4	50.4	0.2021	0.1805	10.7
Room7	211.4	106.5	0.3406	0.2870	15.7

6. Conclusion

In this paper, we propose a method of compensation for low-cost or poorly calibrated IMU based on visual-aided algorithm, which does not require external professional calibration equipment. The visual and inertial information are fused by the coarse-to-fine optimization process. The results in the time domain are further optimized without considering the high-frequency noise in the frequency domain. Our simulations and experimental results in real-world scenarios demonstrate that the calibration approach is effective and can improve the performance of low-cost or poorly calibrated MEMS IMUs in the VIO fusion algorithm. There are advantages in large-scale robot positioning platform deployment and hardware costs.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

This research was partially funded by the Key Research Program of Frontier Science, CAS (Grant No. ZDBS-LY-JSC028) and the National Natural Science Foundation of China (Grant No. 61971399). X Z, Y L, and L X proposed the idea and carried out the programming and experiments. Y L wrote the draft. Z L, P C, and X Z edited and revised the paper. The authors would like to thank P Z and H Z for assistance with the experimental operation and analysis. All authors have read and approved the final paper.

ORCID iDs

Yunfei Liu https://orcid.org/0000-0001-6170-7067
Haifeng Zhang https://orcid.org/0000-0003-2077-9253

References

[1] Qin T, Zheng Y, Chen T, Chen Y and Su Q 2021 RoadMap: a light-weight semantic map for visual localization towards autonomous driving *IEEE Int. Conf. on Robotics and Automation (ICRA)* pp 11248–54

[2] Apple *Apple ARKit* (available at: https://developer.apple.com/arkit/)

[3] Facebook *Oculus VR* (available at: www.oculus.com/)

[4] Brossard M, Barrau A and Bonnabel S 2019 RINS-W: robust inertial navigation system on wheels *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)* pp 2068–75

[5] Herath S, Irandoust S, Chen B, Qian Y, Kim P and Furukawa Y 2021 Fusion-DHL: WiFi, IMU, and floorplan fusion for dense history of locations in indoor environments *IEEE Int. Conf. on Robotics and Automation (ICRA)* pp 5677–83

[6] Miyagusuku R, Yamashita A and Asama H 2018 Data information fusion from multiple access points for WiFi-based self-localization *IEEE Robot. Autom. Lett.* **4** 269–76

[7] Dinh T, Duong N S and Sandrasegaran K 2020 Smartphone-based indoor positioning using BLE iBeacon and reliable lightweight fingerprint map *IEEE Sens. J.* **20** 10283–94

[8] Zheng S, Li Z, Liu Y, Zhang H and Zou X 2022 An optimization-based UWB-IMU fusion framework for UGV *IEEE Sens. J.* **22** 4369–77

[9] Li Z, Yang A, Lv H, Feng L and Song W 2017 Fusion of visible light indoor positioning and inertial navigation based on particle filter *IEEE Photon. J.* **9** 1–13

[10] Ahmad N, Ghazilla R, Khairi N M and Kasi V 2013 Reviews on various inertial measurement unit (IMU) sensor applications *Int. J. Signal Process. Syst.* **1** 256–62

[11] Yang Y, Geneva P, Zuo X and Huang G 2020 Online IMU intrinsic calibration: is it necessary? *Robotics: Science and Systems* (Corvalis, OR: Robotics Science and Systems)

[12] Mur-Artal R, Montiel J M M and Tardós J D 2015 ORB-SLAM: a versatile and accurate monocular SLAM system *IEEE Trans. Robot.* **31** 1147–63

[13] Clark R, Wang S, Wen H, Markham A and Trigoni N 2017 VINet: visual-inertial odometry as a sequence-to-sequence learning problem *Proc. AAAI Conf. on Artificial Intelligence* vol 31 p 1

[14] Li M and Mourikis A I 2012 Improving the accuracy of EKF-based visual-inertial odometry *IEEE Int. Conf. on Robotics and Automation* pp 828–35

[15] Leutenegger S, Lynen S, Bosse M, Siegwart R and Furgale P 2014 Keyframe-based visual–inertial odometry using nonlinear optimization *Int. J. Robot. Res.* **34** 314–34

[16] Mourikis A I and Roumeliotis S I 2007 A multi-state constraint Kalman filter for vision-aided inertial navigation *IEEE Int. Conf. on Robotics and Automation* pp 3565–72

[17] Mur-Artal R and Tardós J D 2017 Visual-inertial monocular SLAM with map reuse *IEEE Robot. Autom. Lett.* **2** 796–803

[18] Qin T, Li P and Shen S 2018 VINS-mono: a robust and versatile monocular visual-inertial state estimator *IEEE Trans. Robot.* **34** 1004–20

[19] Bloesch M, Omari S, Hutter M and Siegwart R 2015 Robust visual inertial odometry using a direct EKF-based approach *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)* pp 298–304

[20] Qin T, Pan J, Cao S and Shen S 2019 A general optimization-based framework for local odometry estimation with multiple sensors (arXiv:1901.03638)

[21] Huang G 2019 Visual-inertial navigation: a concise review *Int. Conf. on Robotics and Automation (ICRA)* pp 9572–82

[22] Liu Y, Li Z, Zheng S, Cai P and Zou X 2022 An evaluation of MEMS-IMU performance on the absolute trajectory error of visual-inertial navigation system *Micromachines* **13** 23

[23] InvenSense *MPU6050* (available at: https://invensense.tdk.com/wp-content/uploads/2015/02/MPU-6000-Register-Map1.pdf)

[24] Glueck M, Buhmann A and Manoli Y 2012 Autocalibration of MEMS accelerometers *IEEE Int. Instrumentation and Measurement Technology Conf. Proc.* pp 1788–93

[25] Glueck M, Oshinubi D, Schopp P and Manoli Y 2014 Real-time autocalibration of MEMS accelerometers *IEEE Trans. Instrum. Meas.* **63** 96–105

[26] Ye L, Guo Y and Su S W 2017 An efficient autocalibration method for triaxial accelerometer *IEEE Trans. Instrum. Meas.* **66** 2380–90

[27] Qureshi U and Golnaraghi F 2017 An algorithm for the in-field calibration of a MEMS IMU *IEEE Sens. J.* **17** 7479–86

[28] Tedaldi D, Pretto A and Menegatti E 2014 A robust and easy to implement method for IMU calibration without external equipments *IEEE Int. Conf. on Robotics and Automation (ICRA)* pp 3042–9

[29] Li Y, Niu X, Zhang Q, Zhang H and Shi C 2012 An *in situ* hand calibration method using a pseudo-observation scheme for low-end inertial measurement units *Meas. Sci. Technol.* **23** 105104

[30] Xiao Y, Ruan X, Chai J, Zhang X and Zhu X 2019 Online IMU self-calibration for visual-inertial systems *Sensors* **19** 7

[31] Rehder J, Nikolic J, Schneider T, Hinzmann T and Siegwart R, 2016 Extending kalibr: calibrating the extrinsics of multiple IMUs and of individual axes *IEEE Int. Conf. on Robotics and Automation (ICRA)* pp 4304–11

[32] Syed Z F, Aggarwal P, Goodall C, Niu X and El-Sheimy N 2007 A new multi-position calibration method for MEMS inertial navigation systems *Meas. Sci. Technol.* **18** 1897

[33] Li M, Yu H, Zheng X and Mourikis A I 2014 High-fidelity sensor modeling and self-calibration in vision-aided inertial navigation *IEEE Int. Conf. on Robotics and Automation (ICRA)* IEEE pp 409–16

[34] Agarwal S and Mierle K, and Others Ceres solver (available at: http://ceres-solver.org)

[35] Mustaniemi J, Kannala J, Sarkka S, Matas J and Heikkila J 2017 Inertial-based scale estimation for structure from motion on mobile devices *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)* pp 4394–401

[36] Forster C, Carlone L, Dellaert F and Scaramuzza D 2017 On-manifold preintegration for real-time visual-inertial odometry *IEEE Trans. Robot.* **33** 1–21

[37] Maye J, Furgale P and Siegwart R 2013 Self-supervised calibration for robotic systems *IEEE Intelligent Vehicles Symp. (IV)* pp 473–80

[38] AnalogDevice *ADIS16490* (available at: www.analog.com/media/en/technical-documentation/data-sheets/adis16490.pdf)

[39] Intel *RealSenseD435i* (available at: www.intelrealsense.com/wp-content/uploads/2022/03/Intel-RealSense-D400-Series-Datasheet-March-2022.pdf)