Real-Time Temporal and Rotational Calibration of Heterogeneous Sensors Using Motion Correlation Analysis

Kejie Qiu, *Member, IEEE*, Tong Qin, *Graduate Student Member, IEEE*, Jie Pan, Siqi Liu, and Shaojie Shen, *Member, IEEE*

***Abstract*—Accurate and robust calibration is crucial to a multisensor fusion-based system. The calibration of heterogeneous sensors is particularly challenging because of the huge difference of the captured sensor data. On the other hand, many calibration approaches ignore temporal calibration that is in fact as important as spatial calibration. In this article, we focus on the temporal calibration of heterogeneous sensors, and the corresponding extrinsic rotation is also derived. Most existing methods are specialized for a certain sensor combination, such as an inertial measurement unit (IMU) camera or a camera-Lidar system. However, heterogeneous multisensor fusion is a tendency in the robotics area, so a unified calibration method is desired. To this end, we leverage the 3-D rotational motion feature for calibration, and auxiliary calibration boards are not needed since multiple odometry methods are available to capture 3-D sensor motion. Using a high-frequency IMU as the calibration reference, an IMU-centric scheme is designed to achieve a unified framework that adapts to various target sensors that can independently estimate 3-D rotational motion. By combining independent IMU-centric calibration pairs, an arbitrary pair of sensors can also be calibrated using the same reference IMU. Due to a novel 3-D motion correlation quantification and analysis mechanism, the temporal offset can be first estimated in real time. Given temporally aligned sensor motion, the extrinsic rotation can be derived in closed-form in the same 3-D motion correlation mechanism. Experimental results of certain sensor combinations show the accuracy and robustness of the proposed method through comparison with state-of-the-art calibration approaches, and the calibration result of a heterogeneous multisensor set demonstrates the scalability and versatility of our method.**

***Index Terms*—Calibration and identification, heterogeneous sensor calibration, probability and statistical methods, sensor fusion.**

Manuscript received March 8, 2020; revised July 22, 2020; accepted September 15, 2020. Date of publication November 25, 2020; date of current version April 2, 2021. This work was supported in part by the Hong Kong Research Grants Council Early Career Scheme under Project 26201616, and in part by the HKUST-DJI Joint Innovation Laboratory. This article was recommended for publication by Associate Editor L. Carlone and Editor P. Robuffo Giordano upon evaluation of the reviewers' comments. *(Corresponding author: Kejie Qiu.)*

Kejie Qiu is with the Alibaba A.I. Labs, Hangzhou 310000, China (e-mail: zjszqkj@gmail.com).

Tong Qin, Jie Pan, Siqi Liu, and Shaojie Shen are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: qintonguav@gmail.com; jpanai@connect.ust.hk; sliubq@connect.ust.hk; eeshaojie@ust.hk).

This article has supplementary downloadable material available at https://ieeexplore.ieee.org, provided by the authors.

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRO.2020.3033698

I. Introduction

MULTISENSOR fusion is fundamental to various robotic applications based on accurate environment perception, such as simultaneous localization and mapping (SLAM) [1] and dynamic environment perception [2], [3]. Different sensors can complement each other and the overall perception capability will be significantly improved with sensor fusion. For example, inertial measurement units (IMUs) feature a high update rate but noisy and drifting sensor data, cameras have high-resolution passive perception but suffer from image blur and scale ambiguity, while Lidars have accurate depth perception ability but limited horizontal resolution and field of view (FOV). To realize valid and robust sensor fusion, sensor data synchronization of different sensors are crucial to such a fusion-based system. In order to achieve high-precision synchronization and calibration, we leverage the high update rate of the IMU and design an IMU-centric calibration scheme, as shown in Fig. 1. Using an IMU as the common calibration reference, all the target sensors can be calibrated with respect to the central IMU, and arbitrary two sensors using the same reference IMU can also be calibrated accordingly.

Many sensor fusion methods assume that the timestamps of different sensors are precisely aligned [1], [4]–[6], which in fact can only be guaranteed by strict hardware synchronization. But for most low-cost and self-built sensor sets, hardware synchronization is not available. In practice, the timestamps of sensor data will be affected by different clocks, triggering mechanisms, transmission delays, data jam, jitter, skew. As [7] points out, correction of the aforementioned factors is called "clock synchronization," while in this article, we will focus on "temporal calibration," namely, the process of determining constant offsets between measurement instants and timestamps. Thus, accurate temporal calibration is the first prerequisite for valid sensor fusion. Most calibration methods are particularly designed for a certain sensor combination, such as a camera-IMU system or a camera-Lidar system. However, multisensor fusion with heterogeneous sensors is common in today's application scenarios, such as autonomous driving. Moreover, new sensors may emerge in the future. Thus, a unified, targetless, real-time, and high-precision calibration scheme for multiple heterogeneous sensors is desirable. In this article, we focus on temporal calibration of heterogeneous sensors based on 3-D motion correlation analysis.

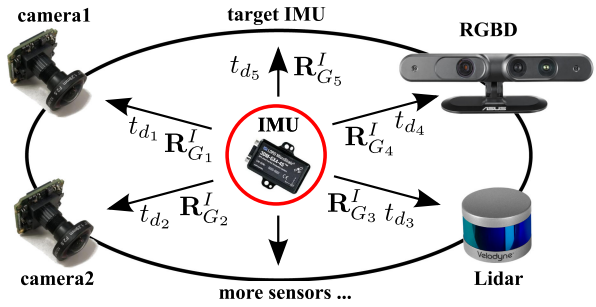


Fig. 1. IMU-centric temporal (temporal offset t_d) and rotational misalignment (extrinsic rotation $\mathbf{R}^I_G$) calibration scheme. For a rigidly connected sensor set, arbitrary two sensors using the same reference IMU can also be calibrated by combining corresponding IMU-centric calibration pairs.

The proposed approach also allows estimating one of the key extrinsic parameters, the extrinsic rotation, using the same 3-D correlation analysis. Our methodology applies to any sensors that can independently estimate 3-D rotational ego-motion.

Most of the existing calibration algorithms belong to the category of the optimization-based method: the goal is to fit/align the sensor data of different sensors, and the calibration problem can be formulated as a state estimation problem and be solved by filter-based or optimization-based solutions. There is another category leveraging motion correlation analysis to calibrate sensors like that reported in [8] and [9]. In fact, correlation analysis is the most standard way to estimate the time-shift between two signals [10], [11]. A possible weakness of the correlation-based method is estimation accuracy. For example, the results of [8] need to be further optimized through an optimization-based method for practical use. Inspired by [8] and [9], we extend the 1-D correlation analysis to 3-D correlation analysis to make full use of the 3-D motion, and design a rate balance filter to balance the update rate difference between the central IMU and the target sensor, resulting in higher calibration accuracy and robustness, which are comparable with optimization-based methods.

As known, every exteroceptive sensor perceives part of the environment in terms of particular physical features, which can be luminous intensities, textures, edges, depths, spectral reflectance, etc. The measured physical quantities are different from one sensor to another, even when two sensors can measure the same quantities, the specific sensing FOV and sensing area can be different. If we make use of the raw sensor data for calibration, accurate sensor data association is the biggest problem. For example, the traditional stereo camera calibration method relies on the epipolar constraint or the reprojection error in the overlapped sensing area for extrinsic calibration [12], which means it cannot handle the case in which the two cameras have no overlapped sensing area. Instead, we utilize the most common physical quantity, ego-motion, which is shared by all the sensors in the same sensor set for heterogeneous sensor calibration, leading to a unified calibration style.

Nowadays, many odometry and SLAM algorithms have been developed to capture the 3-D ego-motion with a particular exteroceptive sensor, such as ORB-SLAM [13] or DSO [14] for monocular cameras, LOAM [15] for Lidars, and RGBD-SLAM [16] for RGBD sensors. While an IMU as a proprioceptive sensor can directly measure the 3-D ego-motion features with angular velocity and linear acceleration. If we extract the motion features as independent signals from different motion estimation methods, auxiliary calibration boards are not needed, and the corresponding temporal offset can be first estimated through a well designed 3-D correlation analysis, which is invariant to multiple geometry transformations. It is a universal fact that the maximum motion correlation observed by two sensors will be obtained with sufficient motion excitation and accurate temporal alignment. After that, the derivation of the extrinsic rotation is similar to a more general shape alignment method, Procrustes analysis [17], [18]. Given temporally aligned motion data, the extrinsic rotation can be further derived in closed-form through eigen direction analysis with the same 3-D correlation analysis mechanism. One advantage of using motion correlation is that the temporal calibration can be applied prior to the extrinsic rotation calibration. Because temporal offset and extrinsic rotation are two uncorrelated quantities, and jointly estimating uncorrelated quantities may potentially impair results as [8] points out. While optimization-based algorithms jointly estimate all states simultaneously since both temporal and spatial misalignments contribute to the fitting error. Also, the proposed method has a much larger estimation range of the temporal offset compared to optimization-based methods. Another advantage is that the extrinsic rotation can be quickly derived in closed-form based on the 3-D correlation analysis used for temporal calibration, which means an initial guess for extrinsic rotation estimation is not needed. While most optimization-based methods rely on initial guesses for accurate convergence.

To the best of the authors' knowledge, this work is the first unified solution to real-time temporal offset and extrinsic rotation estimation of heterogeneous multisensors, without initial guess and extra auxiliary calibration boards. We identify our contributions as follows.

1) We propose a unified, real-time temporal misalignment calibration method for heterogeneous multisensor combinations using robust 3-D motion correlation analysis.
2) We derive a closed-form solution to extrinsic rotation calibration based on the temporal calibration results in the same 3-D correlation analysis mechanism.
3) We show the calibration accuracy and robustness of our method through comparison against ground truth and state-of-the-art calibration approaches with multiple sensor combinations.

The remainder of this article is structured as follows. Section II introduces the relevant work on temporal and spatial calibration between different sensors. In Section III, the problem formulation and system pipeline are illustrated. The 3-D motion correlation evaluation and the key methodology are introduced in Section IV. Section V shows the simulation results of the proposed method using the EuRoC datasets [19]. Section VI gives out the real experimental results and the comparison with the results of other calibration methods. Section VII concludes this article.

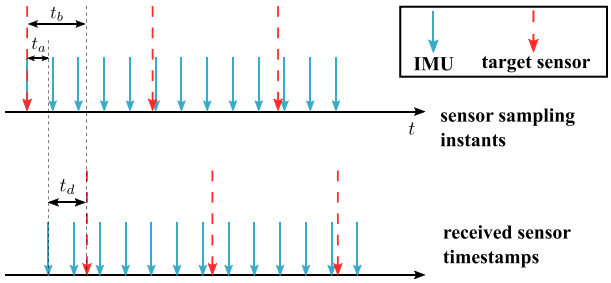


Fig. 2. Temporal misalignment illustration. The difference of sensor latencies leads to temporal offset t_d, which directly determines sensor data alignment.

II. Related Work

Sensor latency is a common issue of real-time applications, and the latency difference of different sensors leads to temporal offset, as shown in Fig. 2. Most temporal calibration methods focus on estimating time offset instead of sensor latency since the misalignment of sensor data will dramatically impact sensor fusion quality. Some sensor fusion-based work assumes that the delay is perfectly known [20]–[23] or the time offset is roughly known [24], [25] in advance.

Recently, a variety of sensor calibration methods have been developed: from sole temporal/spatial calibration [26] to joint calibration [8], [27], from offline calibration [28] to real-time calibration [27], and from closed-form solution [26] to iterative optimization [29]. Whereas, closed-form solutions are not available for temporal calibration. Obviously, real-time calibration methods with a targetless style are more attractive because of their strong adaptiveness and flexibility. Optimization-based solutions based on iterative minimization of nonlinear cost functions are more precise but more computationally expensive, and they require good initial guesses for accurate convergence.

Tungadi and Kleeman propose to estimate the time offset between a Lidar and the wheel odometry of a mobile robot by computing the phase shift of a periodic motion [30]. But spatial alignment of the measurements is required in this method. For IMU-camera calibration, Fleps *et al.* [29] propose to model the sensor trajectory as B-spline and jointly optimizes the control points of the B-spline and the spatial registration parameters. The widely used toolbox Kalibr, proposed by Furgale [28], simultaneously estimates time offset and camera motion, as well as extrinsic rotation and translation between a camera and an IMU with maximum-likelihood estimation using continuous batch optimization, the uncertainty of the estimated offset is also provided. It was then extended for general spatiotemporal calibration in multisensor systems such as Lidar-camera calibration, answering the question of the applicability to synchronization schemes other than hardware synchronization [7]. Furthermore, Rehder *et al.* [9] extend Kalibr to the extrinsic calibration of multiple IMUs. Both of these employ B-spline to parameterize the sensor motion for smooth angular acceleration representation.

The most standard way to estimate the time-shift between two signals is to detect the peak of the cross-correlation between them [10], [11]. For example, Mair *et al.* [8] propose an initialization approach for temporal and spatial registration between an IMU and a camera. It first estimates the time offset using cross-correlation or phase congruency, which is independent of the spatial alignment. Given the estimated time offset, the closed-form rotation alignment between an IMU and a camera is estimated using a modified hand-eye calibration method. However, only the absolute rotational velocity (1-D) is extracted for cross-correlation and phase analysis. It suffers from limited estimation accuracy using 1-D cross-correlation, and fails to obtain accurate temporal alignment of noisy data using phase congruency. Consequently, it is only used as an initialization approach for filter-based or optimization-based methods. We make full use of the 3-D motion and evaluate the motion correlation with 3-D correlation for more robust and accurate calibration, leading to comparable accuracy with optimization-based methods. In fact, 3-D motion correlation analysis has been used for motion decomposition in dynamic object tracking [31], [32].

Recently, several real-time motion estimation methods based on the tightly coupled visual-inertial system treat the time offset as an additional state to be estimated [27], [33]. The time-varying offsets can be handled, and the uncertainty of the estimated offset can also be modeled in this way. The work in [33] is a filter-based method and the work in [27] is a graph optimization-based approach. Both of these only slightly increase the computational complexity compared to the previous frameworks, without estimating time offset. To make the residual error differentiable with respect to time offset, Li and Mourikis [33] apply first-order approximation on the position and orientation of the sensor set with respect to the estimated time offset, while Qin and Shen [27] assume that each feature point moves at a constant speed on the image plane in a short-time interval. Actually, the first-order approximation [33] is another constant speed assumption of the sensor motion itself during a short-time interval. However, they are specially designed for an IMU-camera system. Also, the convergence ranges of both methods are limited due to the constant speed assumptions, in other words, they require good initial guesses for accurate convergence like other iterative solutions.

As for extrinsic parameters estimation, this problem can be abstracted as point set registration from the perspective of statistical shape analysis. It can be solved by the typical Procrustes alignment method [17], [18], which can even find out scaling and reflection deformation in addition to rigid transformation. And the extrinsic rotation can be efficiently computed using the singular value decomposition (SVD). However, Procrustes alignment assumes the data points are well associated, or in other words, the sensor motion data are well synchronized.

Both [33] and [27] keep estimating the extrinsic translation and rotation during motion tracking. Li and Leung [34] propose a spatial-temporal register model using an unscented Kalman filter for multiple dissimilar sensor fusion. Zhang focuses on calibrating the extrinsic rotation between an IMU and a magnetometer using two closed-form solutions [26] such that the iterative procedures in conventional methods can be eliminated. However, Zhang and Song [26] assume the scale factor and bias of the IMU are precalibrated and the sensors are synchronized.

Kelly and Sukhatme [35] propose time delay iterative closest point to estimate the time offset and extrinsic rotation between a proprioceptive and an exteroceptive sensor. The algorithm iteratively computes the spatial and temporal transformations by aligning curves in a 3-D orientation space.

A variety of extrinsic estimation methods designed for Lidar-camera calibration have been proposed recently [36], [37]. Scaramuzza *et al.* [38] introduce an extrinsic calibration technology of a 3-D Lidar and omnidirectional camera using manual point correspondences selection between the camera and Lidar. A unified spatiotemporal calibration method between monocular cameras and planar Lidars is also proposed [39]. The Lidar-camera calibration case in [7] also uses a 2-D Lidar and requires the environment to be at least partly comprised of planes. Faraz *et al.* [40] estimate Lidar-camera extrinsic parameters using an artificial calibration board, in which the intrinsic parameters of the Lidar are also estimated, while Pandey *et al.* [41] propose a targetless extrinsic calibration approach by maximizing the mutual information between Lidar reflectivity and optical image using the Barzilai–Borwein steepest gradient ascent algorithm. A closely related work based on normalized mutual information is also proposed [42], which can be applied to more general environments since it is not solely based on the reflectivity of Lidar, it instead uses particle swarm optimization, which is not restricted to convex problems and so single-scan calibration is possible. However, all these methods are designed for purely extrinsic calibration and assume the sensor data are well synchronized.

III. Overview

Many sensors can independently measure ego-motion in terms of 6-DoF pose, including 3-D rotation. Even a monocular camera can provide up-to-scale ego-motion estimation using monocular visual odometry. We use $\mathbf{p}_y^x$ and $\mathbf{R}_y^x$ to denote the 3-D translation and rotation, respectively, of frame $(\cdot)^y$ with respect to frame $(\cdot)^x$. While for IMUs, accurate pose estimation cannot be obtained individually since the raw IMU measurements are linear acceleration $\hat{\mathbf{a}}$ with gravity and angular velocity $\hat{\boldsymbol{\omega}}$ in the body frame affected by biases and noises

$$\begin{aligned}\hat{\mathbf{a}}_t &= \mathbf{R}_W^{-1}(\mathbf{a}_t - \mathbf{g}^W) + \mathbf{b}_{a_t} + \mathbf{n}_a \\ \hat{\boldsymbol{\omega}}_t &= \boldsymbol{\omega}_t + \mathbf{b}_{g_t} + \mathbf{n}_g \end{aligned} \tag{1}$$

where $\mathbf{b}_a$ is the acceleration bias and $\mathbf{b}_g$ the gyroscope bias, $\mathbf{n}_a$ and $\mathbf{n}_g$ are additional Gaussian white noises.

We classify various IMU-centric combinations into two categories according to the target sensor: IMU-IMU and IMU-general sensor calibration. Without loss of generality, we select the central IMU and one target sensor for calibration illustration, and arbitrary two sensors using the same reference IMU can be calibrated by combining corresponding IMU-centric calibration pairs. The overview system pipeline is shown in Fig. 3, the motion data of the IMU and the target sensor are processed individually until the motion data accumulated in a period of observation duration d_o is sufficient for motion correlation analysis. As thus, the proposed method is a batch correlation analysis-based calibration. However, once the accumulated motion data is sufficient for calibration, our system can run in a sliding window-based way by dropping the oldest motion pair and adding the latest motion pair.

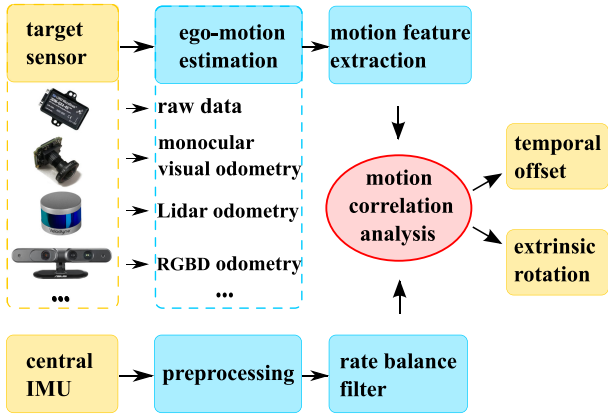


Fig. 3. Overall system pipeline.

As for the frame definitions used in the proposed calibration system, $(\cdot)^I$ and $(\cdot)^G$ are the sensor body frames of the central IMU and the target sensor, respectively. The update rate of the IMU and the target sensor is f_I and f_G Hz, respectively. The odometry of the target sensor is defined with respect to an inertial reference frame, and we define this reference frame as the world frame of the IMU-centric sensor set, denoted as $(\cdot)^W$. Note that this world frame is completely determined by the ego-motion estimation algorithm for the target sensor, it can be anywhere and all the axes of this world frame may not be perpendicular to the horizontal plane. However, the world frame is not needed for IMU-IMU calibration. The relative rotation $\mathbf{R}_G^I$ between frame $(\cdot)^I$ and frame $(\cdot)^G$ denote the key extrinsic rotation, it is an unknown but a constant term in a rigidly connected sensor set.

In this article, we regard the key temporal offset t_d as a constant but unknown value during the observation duration, and extend the IMU-camera time offset definition in our previous work [27] to general target sensors,

$$t_{\text{IMU}} = t_{\text{target}} + t_d. \tag{2}$$

The time offset is the amount of time by which we should shift the timestamps of the target sensor so that the target sensor and central IMU data streams become temporally consistent. It can be a positive or negative value; if the target sensor sequence has a longer latency than the IMU sequence, t_d is a negative value and vice versa.

IV. Methodology

Correlation analysis is a widely used similarity measurement technology, which is particularly suitable for analyzing time-shifted sequences. Instead of performing 1-D correlation evaluation (correlation coefficient) on extracted 1-D motion feature such as absolute angular velocity [8], we directly measure the 3-D correlation on raw 3-D motion, such that high-precision temporal offset estimation and closed-form extrinsic rotation estimation can be achieved. A robust correlation measurement

between two multivariable random processes is trace correlation [43]. It is also used for motion decomposition and scale estimation of a dynamic object [32], in which the metric scale of a dynamic object is optimized by minimizing the correlation between the object motion and the subject motion, while in this article, we propose to calibrate the temporal offset and extrinsic rotation of two sensors by maximizing the corresponding motion correlation.

A. Canonical Correlation Analysis (CCA) and Trace Correlation

Given two random vectors $\mathbf{x} = [x_1, x_2, x_3]^T$ and $\mathbf{y} = [y_1, y_2, y_3]^T$, the cross-covariance and autocovariance can be estimated from the sample data of $\mathbf{x}$ and $\mathbf{y}$ with time shift t

$$\Sigma_{\mathbf{xy}}(t) \approx \frac{1}{N_o - 1} \sum_{n=0}^{N_o-1} (\mathbf{x}_n(0) - \hat{\mathbf{x}}(0))(\mathbf{y}_n(t) - \hat{\mathbf{y}}(t))^T$$

$$\Sigma_{\mathbf{xx}}(t) \approx \frac{1}{N_o - 1} \sum_{n=0}^{N_o-1} (\mathbf{x}_n(0) - \hat{\mathbf{x}}(0))(\mathbf{x}_n(t) - \hat{\mathbf{x}}(t))^T$$

$$\hat{\mathbf{x}}(t) = \frac{1}{N_o} \sum_{n=0}^{N_o-1} \mathbf{x}_n(t), \hat{\mathbf{y}}(t) = \frac{1}{N_o} \sum_{n=0}^{N_o-1} \mathbf{y}_n(t) \tag{3}$$

where the key data association is determined by the time shift t, we will denote $\Sigma_{\mathbf{xy}}(t)$ as $\Sigma_{\mathbf{xy}}$ for simplicity.

The objective of CCA [44] is to find linear combination vector pairs of $\mathbf{a}$ and $\mathbf{b}$ such that the correlation coefficient $\mathrm{Corr}(\mathbf{a}^T\mathbf{x}, \mathbf{b}^T\mathbf{y})$ is maximized

$$\mathbf{a}_i, \mathbf{b}_i = \arg\max_{\mathbf{a},\mathbf{b}} \frac{\mathbf{a}^T \Sigma_{\mathbf{xy}} \mathbf{b}}{\sqrt{\mathbf{a}^T \Sigma_{\mathbf{xx}} \mathbf{a}} \sqrt{\mathbf{b}^T \Sigma_{\mathbf{yy}} \mathbf{b}}}, i \in \{1, 2, 3\}. \tag{4}$$

The random variables $\eta = \mathbf{a}_1^T\mathbf{x}$ and $\phi = \mathbf{b}_1^T\mathbf{y}$ are the first pair of canonical variables. The second pair of canonical variables is derived with the constraints of $\mathbf{a}_2 \perp \mathbf{a}_1, \mathbf{b}_2 \perp \mathbf{b}_1$, and the third pair of canonical variables is derived with the constraints of $\mathbf{a}_3 \perp \mathbf{a}_j, \mathbf{b}_3 \perp \mathbf{b}_j, j \in \{1, 2\}$.

In fact, the three pairs of canonical variables can be calculated in closed-form through eigenvalue decomposition of the matrix $\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}$

$$\begin{aligned} \Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}} &= Q\Lambda Q^{-1} \\ &= \begin{bmatrix} | & | & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix} \begin{bmatrix} - & \mathbf{a}_1^T & - \\ - & \mathbf{a}_2^T & - \\ - & \mathbf{a}_3^T & - \end{bmatrix} \end{aligned} \tag{5}$$

where $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$, and $\mathbf{b}_i$ is collinear with $\Sigma_{\mathbf{y}}^{-1}\Sigma_{\mathbf{yx}}\mathbf{a}_i$

$$\mathbf{b}_i \propto \Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\mathbf{a}_i, i \in \{1, 2, 3\}. \tag{6}$$

The three canonical correlation coefficients are

$$\rho_i = \mathrm{Corr}(\mathbf{a}_i^T\mathbf{x}, \mathbf{b_i}^T\mathbf{y}), i \in \{1, 2, 3\}. \tag{7}$$

Trace correlation between $\mathbf{x}$ and $\mathbf{y}$ is defined as the mean square root of the squares of canonical correlation coefficients

$$\bar{r}(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{1}{3}\sum_{i=1}^{3} \rho_i^2} = \sqrt{\left(\frac{1}{3}\,\mathrm{Tr}(\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}})\right)}. \tag{8}$$

Obviously, $\bar{r}(\mathbf{x}, \mathbf{y}) \in [0, 1]$, and more derivation details on trace correlation can be found in Appendix A.

Similar to the normalized correlation coefficient between two random variables, trace correlation is another normalized measurement between two random vectors that is not affected by absolute value or scaling. Due to the eigenvalue decomposition operation of trace correlation calculation and zero-mean normalization operation of covariance calculation, another property of trace correlation is that it is invariant to scaling, rotation, and translation transformations, which will significantly simplify the correlation measurement between two sensor motions represented in different reference frames, making the temporal offset estimation independent with unknown scaling, rotation, and translation. In other words, the temporal calibration is decoupled from other state estimations. However, for the optimization-based methods, they all need to be jointly estimated. We summarize the multiple geometry transformation invariance properties of trace correlation as

$$\bar{r}(s_{\mathbf{x}} \cdot \mathbf{R}_{\mathbf{x}}\mathbf{x} + \mathbf{p}_{\mathbf{x}}, s_{\mathbf{y}} \cdot \mathbf{R}_{\mathbf{y}}\mathbf{y} + \mathbf{p}_{\mathbf{y}}) = \bar{r}(\mathbf{x}, \mathbf{y}). \tag{9}$$

B. Motion Feature

The key motion feature we use for motion correlation analysis is 3-D body angular velocity, which can be represented by a 3-D random process. First, we directly make use of the raw data of the central IMU as the key calibration reference. For unevenly distributed IMU samples, a preprocessing step for data uniformization is applied based on bilinear interpolation.

For the target sensor except for IMU, on the other hand, we extract the body angular velocity as

$$\lfloor \boldsymbol{\omega} \rfloor_\times = \mathbf{R}^{W\,T}\dot{\mathbf{R}}^W \tag{10}$$

where the skew-symmetric matrix of angular velocity is denoted as $\lfloor \boldsymbol{\omega} \rfloor_\times$

$$\lfloor \boldsymbol{\omega} \rfloor_\times = \begin{bmatrix} 0 & -\boldsymbol{\omega}_z & \boldsymbol{\omega}_y \\ \boldsymbol{\omega}_z & 0 & -\boldsymbol{\omega}_x \\ -\boldsymbol{\omega}_y & \boldsymbol{\omega}_x & 0 \end{bmatrix}. \tag{11}$$

Thus

$$\begin{aligned} \lfloor \boldsymbol{\omega}_I \rfloor_\times = \mathbf{R}_I^{W\,-1}\dot{\mathbf{R}}_I^W &= (\mathbf{R}_G^W\mathbf{R}_I^G)^{-1}(\dot{\mathbf{R}_G^W}\mathbf{R}_I^G) \\ &= \mathbf{R}_G^I \lfloor \boldsymbol{\omega}_G \rfloor_\times \mathbf{R}_G^{I\,-1} \end{aligned} \tag{12}$$

which means

$$\boldsymbol{\omega}_I = \mathbf{R}_G^I \boldsymbol{\omega}_G. \tag{13}$$

For IMU-centric sensor synchronization, we have the following continuous-time relationship according to the IMU measurement model

$$\hat{\boldsymbol{\omega}}_I = \mathbf{R}_G^I \boldsymbol{\omega}_G + \mathbf{b}_{w_t} + \mathbf{n}_w. \tag{14}$$

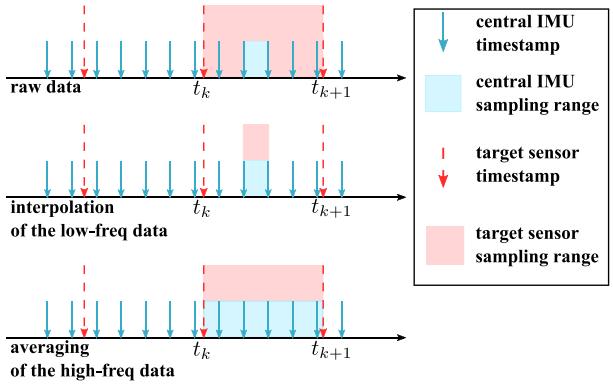


Fig. 4. Motion feature extraction schemes for data association. Top: averaging angular velocity of the target sensor and raw angular velocity of the IMU. Middle: the derived angular velocity of the target sensor from rotation trajectory fitting and raw angular velocity of the IMU. Bottom (rate balance filter): averaging angular velocity of the target sensor and averaging angular velocity of the IMU.

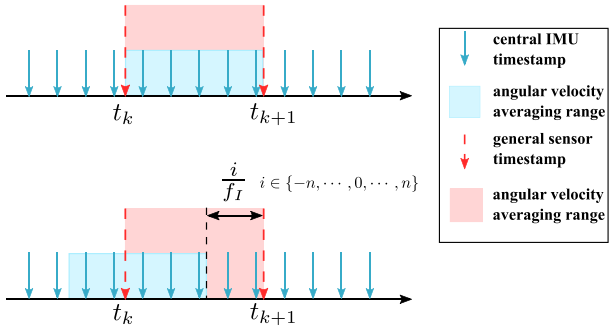


Fig. 5. Enumeration scheme for time-shifted data association.

C. Rate Balance Filter

In practice, the estimated orientations of odometry methods are discretely distributed in frequency f_G, which is much smaller than the IMU sampling frequency f_I. There will be a severe inconsistency problem if we directly use the estimated rotation velocity and the raw IMU rotation velocity for data association. Because the estimated rotation velocity of the target sensor is actually an averaging measurement during Δt ($1/f_G$ seconds from t_k to t_{k+1}), which is longer than the IMU sampling time δt ($1/f_I$ seconds). The averaging body angular velocity of the target sensor can be estimated as

$$\lfloor \boldsymbol{\omega}_G^k \rfloor_\times = \log(\mathbf{R}_{G_{t_k}}^{W\ -1} \mathbf{R}_{G_{t_{k+1}}}^{W}) / \Delta t \tag{15}$$

where $\Delta t = t_{k+1} - t_k$, and $\log(.)$ the logarithmic map from $\mathrm{SO}(3)$ to $\mathrm{so}(3)$.

To deal with this inconsistency, one solution is to fit the target sensor motion using the widely used B-Spline [28], [45] or other interpolation methods, and the rotation velocity can be estimated from the shorter time window δt. But accurate representation using interpolation methods highly depends on the motion model selection and fitting parameter tuning, and it also introduces a large computation burden, as shown in the relevant methods. Thus, instead of interpolating the low-frequency data of the target sensor, we design a rate balance filter, which applies a similar averaging effect on the high-frequency data of the IMU for consistent data association and avoids complex computation. In this way, the motion feature used for data association is actually the averaging angular velocity during $1/f_G$ seconds. The detailed comparison of these two solutions is illustrated in Fig. 4.

To be specific, the delta IMU rotation during the IMU sampling time can be calculated through Lie exponential map from $\mathrm{so}(3)$ to $\mathrm{SO}(3)$

$$\mathbf{R}_{I_t}^{W\ -1} \mathbf{R}_{I_{t+\delta t}}^{W} = \exp(\lfloor \boldsymbol{\omega}_{I_t} \rfloor_\times \delta t). \tag{16}$$

If we consider the duration from t_k to t_{k+1}, the delta IMU rotation can be recovered through integration

$$\begin{aligned} \mathbf{R}_{I_{t_k}}^{W\ -1} \mathbf{R}_{I_{t_{k+1}}}^{W} &= \exp\left(\int_{t_k}^{t_{k+1}} \lfloor \boldsymbol{\omega}_{I_t} \rfloor_\times dt \right) \\ &= \exp(\lfloor \boldsymbol{\omega}_I^k \rfloor_\times \Delta t) \end{aligned} \tag{17}$$

where $\boldsymbol{\omega}_I^k$ is the averaging angular velocity of IMU from t_k to t_{k+1}.

According to (12), we have this equation

$$\left(\int_{t_k}^{t_{k+1}} \lfloor \boldsymbol{\omega}_{I_t} \rfloor_\times dt \right) / \Delta t = \mathbf{R}_G^I \lfloor \boldsymbol{\omega}_G^k \rfloor_\times \mathbf{R}_G^{I\ -1} \tag{18}$$

where $\boldsymbol{\omega}_{I_t} = \hat{\boldsymbol{\omega}}_{I_t} - \mathbf{b}_{g_t}$. During a short period from t_k to t_{k+1}, the gyroscope bias $\mathbf{b}_{g_t}$ can be approximated as a constant term $\mathbf{b}_g^k$.

Define the averaging angular velocity of raw IMU readings as

$$\lfloor \bar{\boldsymbol{\omega}}_I^k \rfloor_\times = \left(\int_{t_k}^{t_{k+1}} \lfloor \hat{\boldsymbol{\omega}}_{I_t} \rfloor_\times dt \right) / \Delta t. \tag{19}$$

Now we can summarize the relationship as

$$\lfloor \bar{\boldsymbol{\omega}}_I^k - \mathbf{b}_g^k \rfloor_\times = \mathbf{R}_G^I \lfloor \boldsymbol{\omega}_G^k \rfloor_\times \mathbf{R}_G^{I\ -1} \tag{20}$$

namely

$$\bar{\boldsymbol{\omega}}_I^k = \mathbf{R}_G^I \boldsymbol{\omega}_G^k + \mathbf{b}_g^k. \tag{21}$$

Since $\bar{\boldsymbol{\omega}}_I^k$ and $\boldsymbol{\omega}_G^k$ satisfy the invariance properties of the trace correlation clarified in (9), both $\mathbf{R}_G^I$ and $\mathbf{b}_g^k$ have no impact on the correlation measurement. Thus, we can measure their temporal correlation directly using trace correlation [see (8)], where $\mathbf{x} = \bar{\boldsymbol{\omega}}_I$ and $\mathbf{y} = \boldsymbol{\omega}_G$, namely $\bar{r}(\bar{\boldsymbol{\omega}}_I, \boldsymbol{\omega}_G)$.

D. Temporal Offset Estimation

Given the measurements for data alignment, the next step is to build the connection of time offset and the corresponding correlation evaluation. Since we only have discrete motion measurements without applying continuous-time interpolation, we solve the optimization problem using time offset enumeration. As shown in Fig. 5, the enumerated $\bar{\boldsymbol{\omega}}_I^k$ of $\boldsymbol{\omega}_G^k$ will be sampled with an IMU sampling interval as the enumeration step, where the start time and the end time for averaging angular velocity

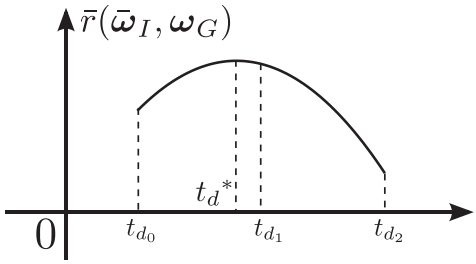


Fig. 6. Time offset refinement using quadratic fitting of the top three sampling points with maximum trace correlation.

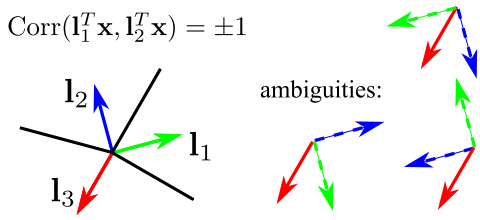


Fig. 7. Sample degenerated case of extrinsic rotation estimation. If $\mathrm{Corr}(\mathbf{l}_1^T\mathbf{x}, \mathbf{l}_2^T\mathbf{x}) = \pm 1$, there exist ambiguous coordinate solutions denoted as the dotted coordinates.

calculation will be shifted by t_d

$$\lfloor \bar{\boldsymbol{\omega}}_I^k(t_d) \rfloor_\times = \left(\int_{t_d+t_k}^{t_d+t_{k+1}} \lfloor \hat{\boldsymbol{\omega}}_{I_t} \rfloor_\times dt \right) / \Delta t \tag{22}$$

where

$$t_d \in \mathcal{T} = \left\{ \frac{i}{f_I}, i = -n, \ldots, 0, \ldots, n \right\}. \tag{23}$$

Since $\Sigma_{\bar{\boldsymbol{\omega}}_I \boldsymbol{\omega}_G}$, $\Sigma_{\bar{\boldsymbol{\omega}}_I \bar{\boldsymbol{\omega}}_I}$, $\Sigma_{\boldsymbol{\omega}_G \boldsymbol{\omega}_I}$, $\Sigma_{\boldsymbol{\omega}_G \boldsymbol{\omega}_G}$, are functions of time offset t_d according to (3), the optimal time offset can be estimated by maximizing the trace correlation

$$t_d{}^* = \arg\max_{t_d \in \mathcal{T}} \bar{r}(\bar{\boldsymbol{\omega}}_I, \boldsymbol{\omega}_G). \tag{24}$$

To further improve the estimation accuracy, we apply quadratic fitting of the top three maximum trace correlation sampling points for estimation refinement, as shown in Fig. 6.

E. Extrinsic Rotation Estimation

Given accurate temporal calibration between two sensors, the relative rotation $\mathbf{R}_G^I$ can be derived under the same 3-D motion correlation evaluation framework. Suppose $\mathbf{d}_i, i \in \{1, 2, 3\}$ are three orthogonal directions in the IMU coordinate space. Obviously, the motion component projected on axis $\mathbf{d}_i$ of $\bar{\boldsymbol{\omega}}_I$ and the motion component projected on axis $\mathbf{R}_G^I \mathbf{d}_i$ of $\boldsymbol{\omega}_G$ are strictly correlated, which means

$$\mathrm{Corr}(\mathbf{d}_i{}^T \bar{\boldsymbol{\omega}}_I, (\mathbf{R}_G^I \mathbf{d}_i)^T \boldsymbol{\omega}_G) = 1, i \in \{1, 2, 3\}. \tag{25}$$

For example, $\mathbf{d}_i$ can be the three axes in the IMU frame $(\cdot)^I$. More specially, they can be the three eigen directions represented in (6). As thus, the relative rotation $\mathbf{R}_G^I$ just corresponds to the matrix $\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}$ ($\mathbf{x} = \bar{\boldsymbol{\omega}}_I, \mathbf{y} = \boldsymbol{\omega}_G$) under the condition of

$$\mathrm{Corr}(\mathbf{a}_i{}^T \bar{\boldsymbol{\omega}}_I, \mathbf{b}_j^T \boldsymbol{\omega}_G) = 1, i = j. \tag{26}$$

According to (7) and (8), this necessary condition equals

$$\bar{r}(\bar{\boldsymbol{\omega}}_I, \boldsymbol{\omega}_G) = 1. \tag{27}$$

In practice, we will soften the necessary condition to some extent for a higher estimation rate, namely

$$\bar{r}(\bar{\boldsymbol{\omega}}_I, \boldsymbol{\omega}_G) > \epsilon_b \tag{28}$$

where $\epsilon_b < 1$ is the bottom threshold for the trace correlation check.

F. Observability Condition

One fatal degenerated motion case of extrinsic rotation is symmetric motion, where the motion components projected on two nonlinear axes are strictly positively or negatively correlated. When symmetric motion occurs, there exist ambiguous coordinate solutions. For example, in Fig. 7, the true coordinate of the target sensor is denoted as the solid coordinate, if the motion projection on the blue axis and green axis are strictly positively correlated or strictly negatively correlated, which means the blue axis and the green axis are swappable, and the coordinate at least has three ambiguous solutions, as shown with the dotted lines. To eliminate possible ambiguous solutions caused by symmetric motion, we further check the following observability condition:

$$\frac{|\lambda_{\max}(\Sigma_{\bar{\boldsymbol{\omega}}_I \bar{\boldsymbol{\omega}}_I})|}{|\lambda_{\min}(\Sigma_{\bar{\boldsymbol{\omega}}_I \bar{\boldsymbol{\omega}}_I})|} < \zeta_u, |\lambda_{\min}(\Sigma_{\bar{\boldsymbol{\omega}}_I \bar{\boldsymbol{\omega}}_I})| > \zeta_b \tag{29}$$

where ζ_u is the upper threshold and ζ_b is the bottom threshold, and $\lambda_{\max}(\Sigma_{\bar{\boldsymbol{\omega}}_I \bar{\boldsymbol{\omega}}_I})$ and $\lambda_{\min}(\Sigma_{\bar{\boldsymbol{\omega}}_I \bar{\boldsymbol{\omega}}_I})$ are the respective maximal and minimal absolute eigenvalue, respectively, of $\Sigma_{\bar{\boldsymbol{\omega}}_I \bar{\boldsymbol{\omega}}_I}$. The detailed derivation regarding the observability condition can be found in Appendix B. The thresholds are set empirically as a tradeoff between the estimation accuracy and estimation rate, which will be detailed in the simulation and experiment sections.

With the necessary condition and observability condition satisfied, the extrinsic rotation matrix can be estimated accordingly with the constraints of $\mathbf{R}\mathbf{R}^T = \mathbf{I}$ and $\det(\mathbf{R}) = 1$

$$\mathbf{R}_G^I = \mathbf{U} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det(\mathbf{U}\mathbf{V}^T) \end{bmatrix} \mathbf{V}^T \tag{30}$$

where $\mathbf{U}$ and $\mathbf{V}$ is the left and right unitary matrix of the SVD of $\Sigma_{\boldsymbol{\omega}_G \boldsymbol{\omega}_G}^{-1} \Sigma_{\boldsymbol{\omega}_G \bar{\boldsymbol{\omega}}_I}$

$$\Sigma_{\boldsymbol{\omega}_G \boldsymbol{\omega}_G}^{-1} \Sigma_{\boldsymbol{\omega}_G \bar{\boldsymbol{\omega}}_I} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T. \tag{31}$$

The extrinsic rotation is a byproduct of temporal offset estimation since it can be computed in closed-form given temporally aligned motion data, only an SVD operation and the observability check is needed. In other words, the increased computational complexity is almost zero for extrinsic rotation estimation in our calibration framework, which further ensures the real-time performance within a multisensor system.

The whole algorithm is summarized as Algorithm **1**.

Algorithm 1: IMU-Centric Sensor Calibration Algorithm.

Input: orientation estimations of the target sensor in world frame $\mathbf{R}^W_{G_{t_k}}$ (f_G Hz), angular velocity data of IMU $\hat{\boldsymbol{\omega}}_{I_t}$ (f_I Hz), time offset enumeration range t_r, motion observation duration d_o.

Output: timestamp offset t_d^*, extrinsic rotation $\mathbf{R}^I_G$

IMU uniformization and upsampling based on bilinear interpolation

$\lfloor \boldsymbol{\omega}^k_G \rfloor_\times = \log(\mathbf{R}^W_{G_{t_k}}{}^{-1}\mathbf{R}^W_{G_{t_{k+1}}})/(t_{k+1}-t_k)$ ▷Extract averaging angular velocity of the target sensor

$\mathcal{T} = \{i/f_I, i = -n, \ldots, 0, \ldots, n\}$ where $n = t_r * f_I$▷Enumeration set of time offset

for $t_d \in \mathcal{T}$ **do**

$\lfloor \bar{\boldsymbol{\omega}}^k_I(t_d) \rfloor_\times = (\int_{t_d+t_k}^{t_d+t_{k+1}} \lfloor \hat{\boldsymbol{\omega}}_{I_t} \rfloor_\times dt)/(t_{k+1}-t_k)$ ▷Extract averaging angular velocity of IMU of t_d

end for

$[\boldsymbol{\omega}^k_G, \bar{\boldsymbol{\omega}}^k_I(t_d)] \rightarrow$ **buffer** ▷Motion data accumulation (old motion data beyond d_o will be removed)

if $\mathrm{duration}(\mathbf{buffer}) < d_o$ **then**

return

endif

for $t_d \in \mathcal{T}$ **do**▷Time offset enumeration

$\Sigma_{\bar{\boldsymbol{\omega}}_I \boldsymbol{\omega}_G}(t_d)$, $\Sigma_{\bar{\boldsymbol{\omega}}_I \bar{\boldsymbol{\omega}}_I}(t_d)$, $\Sigma_{\boldsymbol{\omega}_G \boldsymbol{\omega}_I}(t_d)$, $\Sigma_{\boldsymbol{\omega}_G \boldsymbol{\omega}_G}(t_d)$▷Update covariance of t_d

$\bar{r}(\bar{\boldsymbol{\omega}}_I, \boldsymbol{\omega}_G)|_{t_d} = \sqrt{\frac{1}{3}\,\mathrm{Tr}(\Sigma^{-1}_{\bar{\boldsymbol{\omega}}_I \bar{\boldsymbol{\omega}}_I} \Sigma_{\bar{\boldsymbol{\omega}}_I \boldsymbol{\omega}_G} \Sigma^{-1}_{\boldsymbol{\omega}_G \boldsymbol{\omega}_G} \Sigma_{\boldsymbol{\omega}_G \bar{\boldsymbol{\omega}}_I})}$ ▷ Trace correlation calculation of t_d

endfor

$t_d{}^* = \arg\max_{t_d \in \mathcal{T}} \bar{r}(\bar{\boldsymbol{\omega}}_I, \boldsymbol{\omega}_G)$

if $\bar{r}(\bar{\boldsymbol{\omega}}_I, \boldsymbol{\omega}_G)|_{t_d = t_d{}^*} > \epsilon_b$ **and** $\frac{|\lambda_{\max}(\Sigma_{\bar{\boldsymbol{\omega}}_I \bar{\boldsymbol{\omega}}_I})|}{|\lambda_{\min}(\Sigma_{\bar{\boldsymbol{\omega}}_I \bar{\boldsymbol{\omega}}_I})|} < \zeta_u$ **and** $|\lambda_{\min}(\Sigma_{\bar{\boldsymbol{\omega}}_I \bar{\boldsymbol{\omega}}_I})| > \zeta_b$

$\Sigma^{-1}_{\boldsymbol{\omega}_G \boldsymbol{\omega}_G} \Sigma_{\boldsymbol{\omega}_G \bar{\boldsymbol{\omega}}_I} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ ▷SVD

$$\mathbf{R}^I_G = \mathbf{U} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det(\mathbf{U}\mathbf{V}^T) \end{bmatrix} \mathbf{V}^T$$

return $t_d{}^*$, $\mathbf{R}^I_G$

endif

V. Simulation

Instead of building a completely virtual simulation environment, we perform simulation based on the EuRoC MAV Dataset [19], which contains synchronized stereo images and IMU data with ground truth. The images (Aptina MT9V034 global shutter, WVGA monochrome, 20 FPS) and IMU measurements (ADIS16448, 200 Hz) are strictly synchronized due to the well-designed hardware synchronization scheme in this dataset. We only use images from the left camera as the target sensor, and we can conveniently set the time offset by manually shifting IMU timestamps with a certain value. The whole calibration algorithm runs real-time on a desktop computer with an i7-8700 K CPU. The time offset enumeration range is from -1100 to 1100 ms with a step of 5 ms. The threshold values used during calibration are $\epsilon_b = 0.9$, $\zeta_b = 0.03$, and $\zeta_u = 10$.

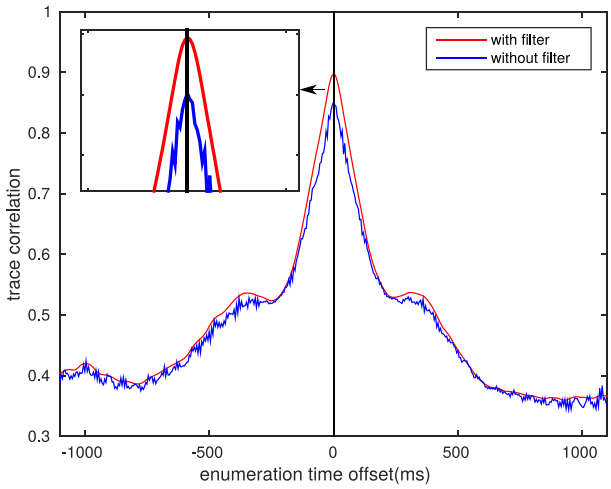


Fig. 8. Trace correlation curve comparison with and without applying the rate balance filter (see Section IV-C).

A. Rate Balance Filter

The trace correlation curve regarding enumerated time offset directly determines the temporal calibration quality, with which the extrinsic rotation is also computed. First, we intuitively compare two kinds of trace correlation curves with and without the rate balance filter. As shown in Fig. 8, the curve with rate balance filter is more smooth and the peak is more distinct. If we focus on the curve peak area, we can notice that the peak of the curve is more symmetric compared with that without applying the filter, which is beneficial for further estimation refinement. In addition, the curve with rate balance filter achieves stronger correlation at the peak, which illustrates the designed rate balance filter results in better motion data association.

B. Temporal Offset Range

One key advantage of the proposed method over optimization-based methods is the large estimation range of the temporal offset. To demonstrate the estimation results with different time offsets, we manually add time shift to the IMU timestamps of Dataset V1_01. Sample trace correlation curves with three different time offset values are shown in Fig. 9, as we can see from the figure, different time offsets only lead to curve shifting of the corresponding trace correlation curves, which means the estimation accuracy and variance of the time offset are irrelevant to the ground truth time offset. Thus, the estimation range of the proposed method solely depends on the enumeration range (-1100–1100 ms), which can be adjusted into a reasonable range according to specific calibration cases. As for the optimization-based methods [27], [33], on the other hand, the time offset estimation range is much smaller because of the constant speed approximation during a short-time interval in their formulations. For example, Qin and Shen[27] converge slowly or even fails to estimate the temporal misalignment when the time offset is larger than 50 ms according to our test. The detailed estimation comparison can be found in Section VII.

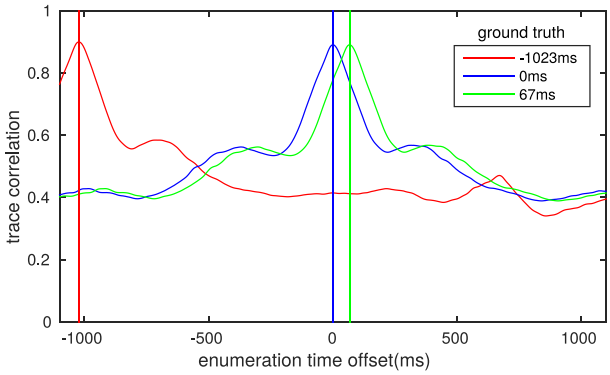


Fig. 9. Time offset simulations and corresponding trace correlation curves of Dataset V1_01. Different time offsets lead to curve shifting of the corresponding trace correlation, in other words, the estimation accuracy and variance of the time offset are irrelevant with the ground truth time offset.

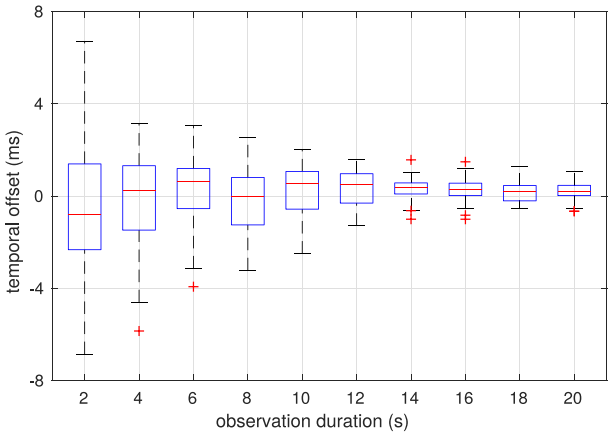


Fig. 10. Boxplot of the temporal offset estimation of Dataset V1_01 (ground truth $t_d = 0$ ms). A longer observation duration d_o leads to better estimation accuracy but a larger estimation delay.

C. Observation Duration

Since we accumulate a period of time of motion observation for sensor calibration, the observation duration is key to calibration accuracy. A longer observation duration is preferred because the motion correlation measurement, which depends on statistical data, can be more accurate; however, too long an observation duration leads to longer estimation delay and violates the IMU bias approximation in a short time (see Section IV-C). Thus, a proper observation duration is desired for effective and efficient calibration. We conduct an experiment on observation duration with dataset V1_01. And we apply a sliding window-based estimation method following Algorithm **1**, in which the window width is the observation duration. The relationship between estimation accuracy and observation duration is shown in Fig. 10; the accuracy increases quickly with longer observation duration within 10 s and then gradually becomes stable.

D. Estimation Accuracy

We then evaluate the estimation accuracy through running the proposed calibration method (observation duration = 8 s) with the raw sensor data without adding a time shift to the IMU timestamps since the estimation accuracy is irrelevant to the specific time offset value. The statistics results including mean value and standard deviation (std) are shown in Table I. We use Z–Y–X Euler angles for rotation representation. We can see that the accuracy of time offset estimation can be within 1.2 ms, and the accuracy of the extrinsic rotation is within 1.8°. As comparison, we show the time offset estimations using 1-D correlation analysis, the accuracy is inferior to that of using the proposed 3-D correlation analysis considering that the ground truth time offset is 0 ms. A boxplot of the comparison is also shown in Fig. 11. Besides that, the one using 1-D correlation cannot directly estimate the extrinsic rotation.

TABLE I
SIMULATION CALIBRATION RESULTS

dataset	corr[1]	time offset[2]		yaw	pitch	roll
		mean(ms)	std(ms)	std(°)		
MH_01	1D	-1.233	0.600	-	-	-
	3D	-0.029	0.739	0.083	0.267	0.285
MH_03	1D	-1.144	1.376	-	-	-
	3D	-0.075	0.598	1.129	0.873	3.745
MH_05	1D	-2.218	2.206	-	-	-
	3D	-0.633	1.221	0.445	1.730	2.998
V1_01	1D	-1.435	2.878	-	-	-
	3D	-0.261	1.227	0.913	1.029	1.484

[1] 1-D: 1-D motion correlation; 3-D: 3-D motion correlation.
[2] The ground truth time offset is 0 ms of EuRoC datasets.

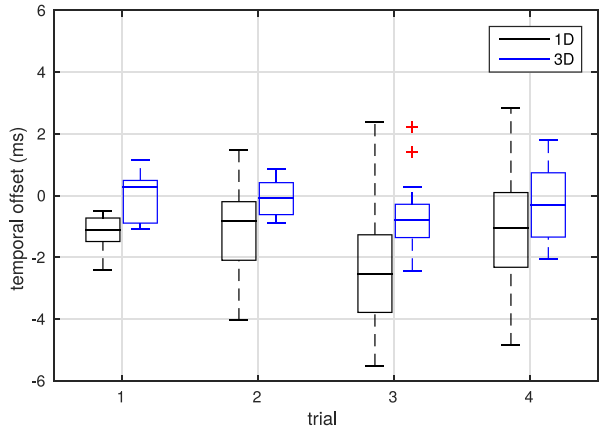


Fig. 11. Temporal offset estimations using 1-D correlation and 3-D trace correlation.

VI. EXPERIMENTS

A. Implementation Details

As shown in Fig. 12, the sensor set used for IMU-IMU calibration consists of a Microstrain 3DM-GX4 IMU that runs at 500 Hz and the internal IMU of an MYNT EYE Standard stereo camera[1] that also runs at 500 Hz. The sensor set used for IMU-camera calibration includes the internal IMU and the

[1] [Online]. Available: https://www.mynteye.com/

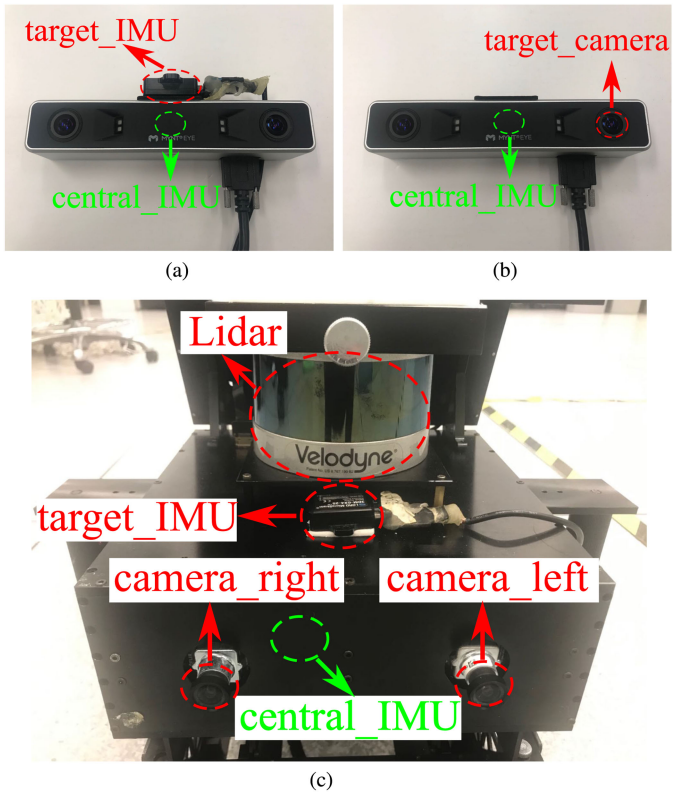


Fig. 12. Sensor combinations for different experiments. (a) IMU-IMU. (b) IMU-camera. (c) Heterogeneous multi-sensor set.

left monocular camera of the MYNT camera, which runs at 30 Hz. Finally, we build a black box that contains two Point Grey Chameleon3 cameras[2] that run at 30 Hz, the IMU inside the DJI A3 flight controller[3] that runs at 400 Hz, the same 3DM-GX4 IMU, and a Velodyne VLP-16Puck LITE Lidar[4] that runs at 20 Hz to present heterogeneous multisensor calibration. To be specific, the IMU inside the DJI A3 and the Velodyne Lidar are used for IMU-Lidar calibration, while an additional Point Grey camera is used for Lidar-IMU-camera calibration. Please note the relatively high-end 3DM-GX4 IMU is only used as the target sensor in the IMU-IMU calibration case, all the central reference IMUs are low-cost ones, including the internal IMUs inside the MYNT camera and DJI A3. We use the Linux OS with ROS[5] as the development middleware. The time offset enumeration range is from -1000 to 1000 ms with a step of 2.5 ms, and the motion observation duration d_o is 4 s for IMU-IMU calibration and 8 s for other IMU-centric calibrations. The threshold values used during calibration are $\epsilon_b = 0.9$, $\zeta_b = 0.015$, and $\zeta_u = 20$.

B. IMU-IMU

For calibration between IMUs, extra ego-motion estimation method for the target sensor is not needed. We compare our method with Extending Kalibr [9], which extends the toolbox Kalibr [28] to determine IMU extrinsics and intrinsics. And we focus on the extrinsic estimations for comparison. Besides direct IMU-IMU calibration, the author in [9] can indirectly derive the IMU-IMU calibration results through one or multiple exteroceptive sensors as the common calibration reference. For fair IMU-IMU calibration comparison, we choose the direct calibration mode of Extending Kalibr, in which the extrinsic translation is not calibrated. And a default third-order B-spline (50 knots per second) is employed in Extending Kalibr, which encodes the angular acceleration as a cubic polynomial. The central IMU and the target IMU we use are shown in Fig. 12(a). We collect five independent datasets by moving the sensor set randomly for 60 s. We then split each dataset into 15 clips of 4 s.

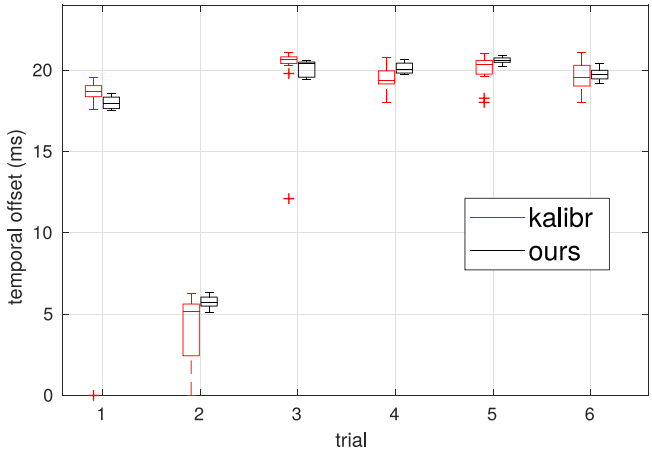


Fig. 13. IMU-IMU temporal offset estimation comparison. Kalibr: [9].

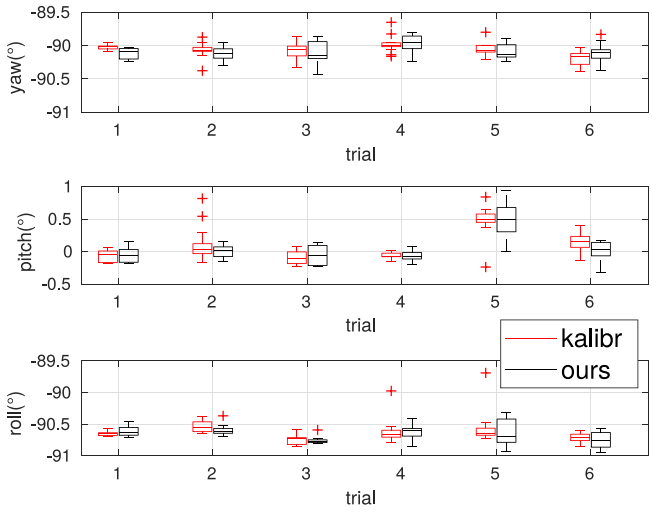


Fig. 14. IMU-IMU extrinsic rotation estimation comparison. Kalibr: [9].

The results of temporal calibration and extrinsic rotation calibration are depicted in Figs. 13 and 14, respectively. Since ground truth is unavailable, we take Kalibr's results as a reference. Our results are quite close to Kalibr's results. We can see that some outliers appear (red "+") for temporal offset estimation using Kalibr, especially in trial 1 and 2, and in trial 3

[2][Online]. Available: https://www.ptgrey.com/chameleon3-usb3-vision-cameras
[3][Online]. Available: https://www.dji.com/a3
[4][Online]. Available: https://velodynelidar.com/vlp-16-lite.html
[5][Online]. Available: http://www.ros.org/

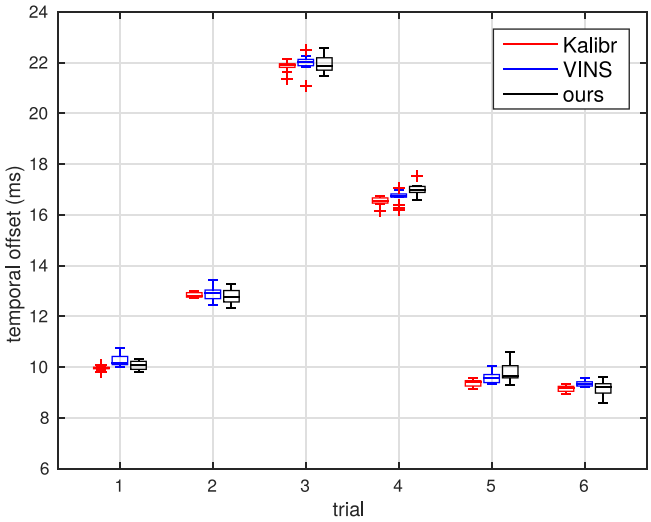


Fig. 15. IMU-camera temporal offset estimation comparison. Kalibr: [28]; VINS: [27].

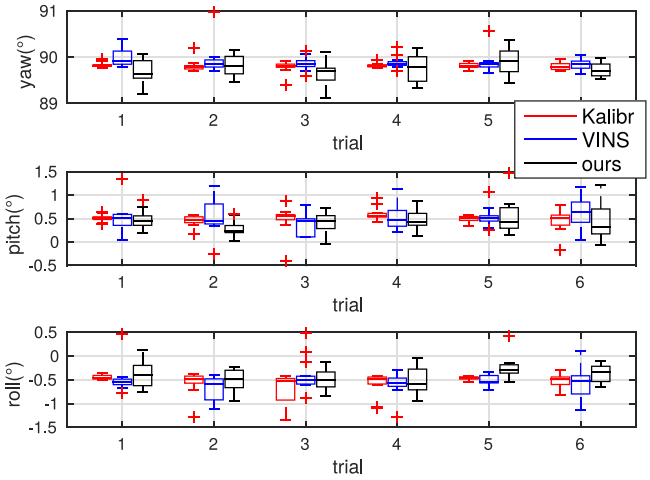


Fig. 16. IMU-camera extrinsic rotation estimation comparison. Kalibr: [28]; VINS: [27].

TABLE II
IMU-CAMERA TEMPORAL OFFSET STATISTICS

trial	Kalibr [28]		VINS [27]		ours	
	mean (ms)	std (ms)	mean	std	mean	std
1	9.9547	0.0594	10.248	0.2300	10.061	0.2073
2	12.836	0.0893	12.896	0.2465	12.794	0.2861
3	21.846	0.1853	21.991	0.3087	21.944	0.3760
4	16.542	0.1514	16.717	0.2531	16.994	0.0884
5	9.3554	0.1325	9.5849	0.2185	9.7908	0.3158
6	9.1498	0.1188	9.3425	0.1079	9.1422	0.2518

(the outlier 434 ms is beyond the axis scope), while our method remains robust for all trials. For the extrinsic rotation calibration, both methods perform steadily with a small variance. Again our method has fewer outlier estimations compared to Kalibr. Thus, for IMU-IMU calibration, the proposed method using angular velocity as the key motion feature outperforms Kalibr in terms of consistency and robustness.

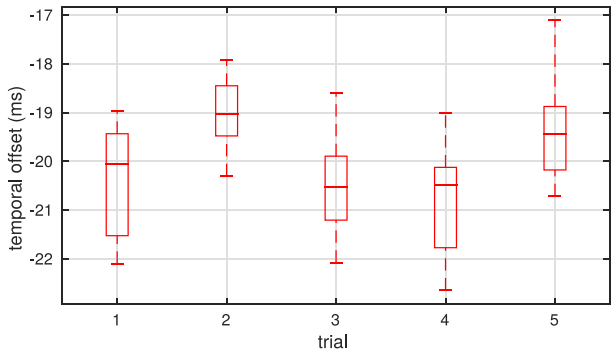


Fig. 17. IMU-Lidar temporal offset estimation.

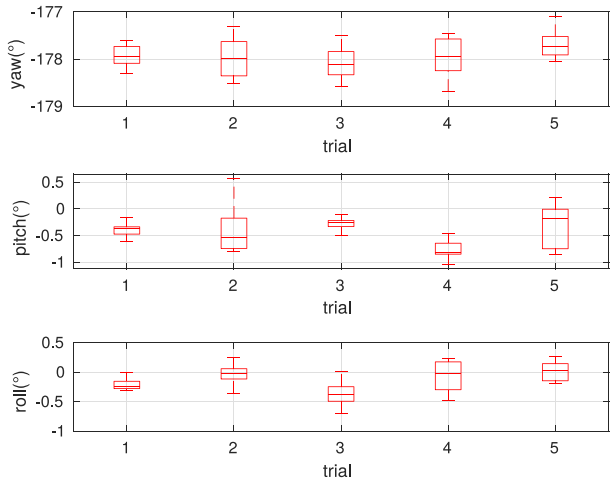


Fig. 18. IMU-Lidar extrinsic rotation estimation.

C. IMU-Monocular Camera

For calibration between an IMU and a monocular camera, we use ORB-SLAM [13] as the ego-motion estimation method for the monocular camera. We compare our method with Spatio-Temporal Initialization [8], Kalibr [28], and VINS [27]. Since Kalibr relies on a calibration board for IMU-camera calibration, a wide-angle camera is used to keep observing the whole calibration board during excitation motion. The central IMU and the target camera we use are shown in Fig. 12(b). We collect six independent datasets by moving the sensor set before a checkerboard for 300 s, then split each dataset into 15 clips of 20 s. Please note the checkerboard is only used by Kalibr for robust camera motion estimation, the proposed method instead utilizes ORB-SLAM as the ego-motion estimation front end for targetless calibration purpose.

The results of temporal calibration and extrinsic rotation calibration are depicted in Figs. 15 and 16, respectively. Since ground truth is unavailable, we take Kalibr's results as a reference. The detailed temporal offset calibration results are shown in Table II. Our results are quite close to the estimations of Kalibr and VINS. For the extrinsic rotation calibration, all the methods perform steadily with a small variance. Thus, for IMU-camera calibration, the proposed method has comparable

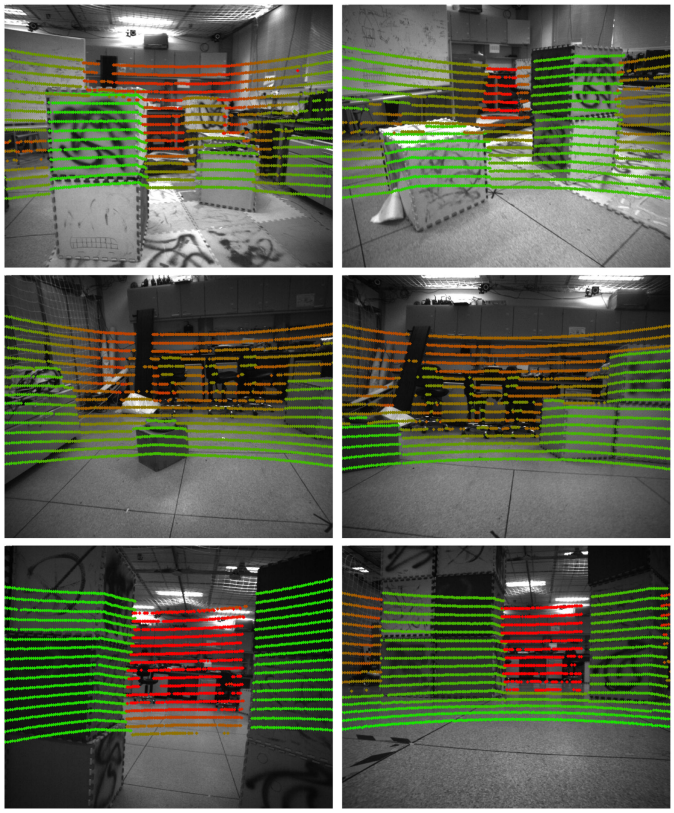


Fig. 19. Visualization of the reprojected Lidar points on the images using the calibrated temporal offset and extrinsic rotation (the extrinsic translation is manually measured in advance). The point color encodes the point depth from the camera center (green to red: near to far). There exist some occlusion points because the Lidar and camera are not concentric.

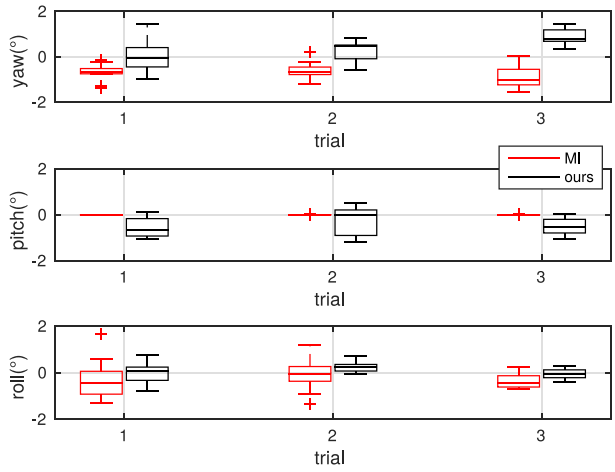


Fig. 20. Lidar-camera extrinsic rotation estimation. MI: mutual information-based method [41].

TABLE III
LIDAR-CAMERA EXTRINSIC ROTATION STATISTICS

trial	yaw std (°)		pitch std (°)		roll std (°)	
	MI*	ours	MI	ours	MI	ours
1	0.372	0.645	0.005	0.412	1.026	0.387
2	0.369	0.475	0.006	0.605	0.816	0.227
3	0.452	0.529	0.011	0.354	1.091	0.195

* MI: mutual information-based method [41].

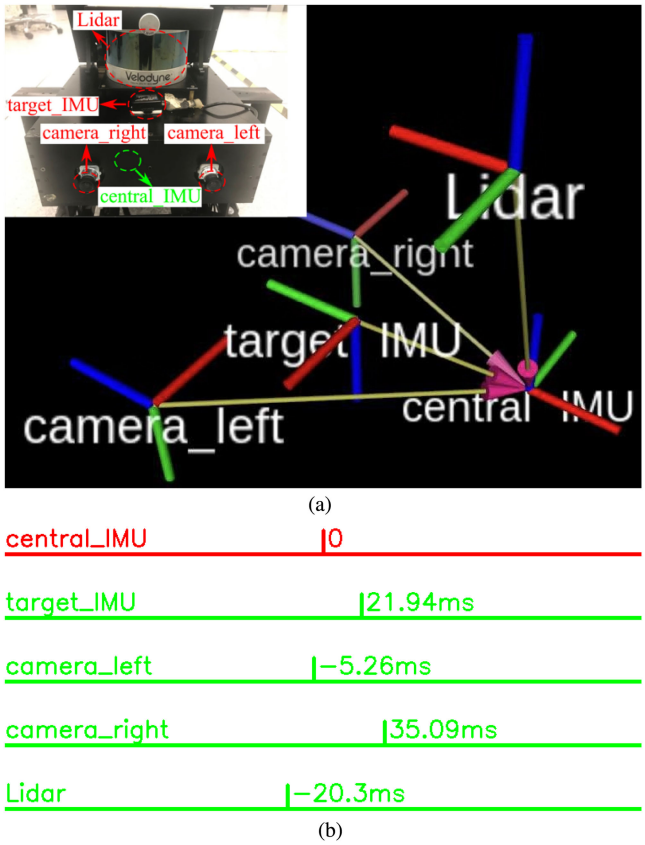


Fig. 21. Overall calibration results of the heterogeneous multisensor set. (a) Extrinsic estimation. (b) Temporal offset.

TABLE IV
TIMING STATISTICS

Target sensor	Module	Time (ms)	Rate (Hz)
target IMU	data preprocessing	0.5	50[1]
	t_d estimation	5	50[1]
	R_G^I estimation	0[2]	50[1]
camera (left/right)	visual odometry	27[3]	30
	data preprocessing	1.2	30
	t_d estimation	7	30
	R_G^I estimation	0[2]	30
Lidar	Lidar odometry	10[4]	20
	data preprocessing	1.8	20
	t_d estimation	6	20
	R_G^I estimation	0[2]	20

[1]The processing rate is downsampled to 50 Hz.
[2]The extrinsic rotation is a byproduct of temporal offset estimation (Sections IV-E and IV-F).
[3]ORB-SLAM [13] is used for visual odometry implementation.
[4]LOAM [15] is used for Lidar odometry implementation.

estimation accuracy with Kalibr [28] and VINS [27], which are both optimization-based methods.

D. IMU-Lidar

For calibration between an IMU and a Lidar, we use LOAM [15] as the ego-motion estimation method for the Lidar. The central IMU and the target Lidar, we use are shown in

TABLE V
ALGORITHM COMPARISON

	sensor set	estimation approach	domain (realization)	t_d	$\mathbf{R}$	$\mathbf{T}$	real time	targetless	t_d range	accuracy
Kalibr [28]	IMU-camera	batch optimization	CT (spline fitting)	•	•	•	-	-	0.01s*	high
Kalibr* [9]	IMU-IMU	1D xcorr maximization	CT (spline fitting)	•	•	-	-	-	-	high
Elmar [8]	IMU-camera	1D xcorr maximization	DT (none)	•	•	-	-	-	-	low
VINS [27]	IMU-camera	batch optimization	CT (linearization)	•	•	•	•	•	0.05s*	high
Li [33]	IMU-camera	MSCKF	CT (linearization)	•	•	•	•	•	0.05s*	high
MI [41]	Lidar-camera	batch optimization	-	-	•	•	-	•	-	high
Ours	**multi-sensors**	**3D** tcorr maximization	DT (RBF)	•	•	-	•	•	$>$ **1s**	high

xcorr: cross-correlation; tcorr: trace correlation; CT: continuous time; DT: discrete time; RBF: rate balance filter; MI: mutual information; •: satisfied; -: not satisfied or not available; spline fitting depends on parameters tuning (basis order and knots density); t_d: temporal offset; $\mathbf{R}$: extrinsic rotation; $\mathbf{T}$: extrinsic translation; Targetless: calibration boards like chessboards not needed. [*]: The convergence range of t_d are not explicitly reported, we only show the maximum estimation range shown in the corresponding papers.

Fig. 12(c). Consider that there are no benchmark datasets or methods available on IMU-Lidar calibration, we prefer to do the experiments on a strictly synchronized IMU and Lidar dataset like the EuRoC dataset. To this end, we build a synchronization circuit triggered by the central IMU to continuously correct the onboard clock inside the Lidar through the GPS port of Velodyne Lidar. This hardware synchronization eliminates the clock drift but a constant time offset. We collect five independent datasets by moving the sensor set for around 60 s, and split each dataset into 15 overlapped clips for evaluation.

The boxplot results of temporal calibration and extrinsic rotation calibration are depicted in Figs. 17 and 18, respectively. As we can see, the time offset is around the level of -20 ms, the standard deviation of the time offset estimation is within 1 ms for each dataset, and the standard deviation of yaw, pitch, and roll estimations are within 0.4°, 0.5°, and 0.3°, respectively. The temporal calibration results are not as stable as those of the IMU-camera calibration may be caused by the motion distortion of the point cloud although LOAM can correct this distortion to some extent.

E. Lidar-Camera

Different from a direct data association between optical images and lidar points for Lidar-camera calibration [41], we instead derive the temporal and extrinsic rotation calibration results from two independent IMU-camera and IMU-Lidar calibrations. And we use ORB-SLAM [13] and LOAM [15] as the ego-motion estimation methods for the camera and Lidar, respectively. The time offset between the camera and Lidar is computed by

$$\begin{aligned} t_{\text{IMU}} &= t_{\text{camera}} + t_{d_1} \\ t_{\text{IMU}} &= t_{\text{Lidar}} + t_{d_2} \\ t_{\text{camera}} &= t_{\text{Lidar}} + t_{d_2} - t_{d_1} \end{aligned} \tag{32}$$

and the extrinsic rotation between the camera and Lidar is

$$\mathbf{R}_L^C = {\mathbf{R}_C^I}^T \mathbf{R}_L^I. \tag{33}$$

Given the temporal offset and extrinsic parameters (the extrinsic translation between the camera and Lidar is manully measured in advance), we can, thus, reproject the Lidar points onto the image plane for vivid visualization and further use. We conduct an experiment using the central IMU, Lidar, and camera_left, as shown in Fig. 12(c), in an indoor environment with boxes inside it. Sample reprojection images are shown in Fig. 19. The depth of the Lidar points are encoded using different colors (near to far: green to red). As we can see, the scene edges are well distinguished by the reprojected Lidar points, which is proof of the estimated temporal offset and extrinsic rotation of our calibration method.

Also, we implement the mutual information-based extrinsic calibration method proposed in [41] for numerical comparison. Consider that this method cannot estimate the time offset, we apply the same time offset compensation obtained by our method for a fair comparison. We collect three independent datasets by moving the sensor set randomly for around 150 s, and split each dataset into 15 clips for evaluation. Since we use Euler angle for illustration, in order to avoid gimbal lock with Euler angle representation, we apply a common transformation on both estimation results for the final Euler angle conversion

$$\mathbf{R}_{\text{ini}}^{-1}\mathbf{R}_L^C \rightarrow (\text{yaw, pitch, roll}), \mathbf{R}_{\text{ini}} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{bmatrix} \tag{34}$$

where $\mathbf{R}_{\text{ini}}$ is the initial guess of the extrinsic rotation.

The extrinsic rotation estimation comparison on the three datasets are shown in Fig. 20, and the standard deviation results are listed in Table III. Note that [41] is an optimization-based method, which highly relies on the initial guess, the corresponding pitch estimations are all almost 0°, which means it has almost no adjustment around the initial guess that is in fact not the ground truth. Our method, on the other hand, does not require initial guess for extrinsic rotation estimation because the closed-form solution is available. The overall performance of the proposed method on indirect Lidar-camera calibration is comparable with the dedicated Lidar-camera calibration method.

F. Heterogeneous Multisensor Calibration

To demonstrate the overall calibration results of heterogeneous multisensors, we use all the sensor data provided by the black box including two IMUs, two cameras, and one Lidar, as shown in Fig. 12(c). To be specific, the internal low-cost IMU is used as the common calibration reference and four independent IMU-centric calibration threads run simultaneously. One screenshot of the temporal offset and extrinsic rotation estimation results are shown in Fig. 21, in which the extrinsic translations

are measured in advance. In fact, the most time-consuming modules are the corresponding odometry methods of the target sensors, the detailed timing statistics are shown in Table IV.

VII. Conclusion

In this article, we proposed an IMU-centric temporal offset and extrinsic rotation estimation algorithm for heterogeneous multisensor calibration. The calibration method features comparable estimation accuracy with optimization-based methods, a much larger temporal offset estimation range and closed-form extrinsic rotation estimation. It can work in natural scenes without auxiliary calibration boards in real time. The algorithm comparison is summarized in Table V. The proposed method outperforms other state-of-the-art methods in terms of versatility, computation efficiency, and large estimation range of the temporal offset. In addition, multiple independent calibration threads using the same IMU as the central reference can work together to calibrate the temporal misalignment and extrinsic rotation between arbitrary two sensors in the same sensor set. The degenerated cases and corresponding observability conditions were studied to further improve the calibration robustness. We validated the estimation accuracy and versatility of the proposed algorithm through simulation on benchmark datasets and extensive real-world experiments.

In the future, the IMU-centric calibration scheme can be extended to directly calibrate two sensors as long as the update rate of one sensor is sufficiently high, and the central IMU is no longer needed. For example, a high-speed camera such as the emerging event camera [46] is a good alternative to the central calibration reference.

Appendix A
CCA and Trace Correlation

Define

$$\mathbf{c} = \Sigma_{\mathbf{xx}}^{1/2}\mathbf{a}, \mathbf{d} = \Sigma_{\mathbf{yy}}^{1/2}\mathbf{b} \tag{35}$$

for (4), so we have

$$\mathrm{Corr}(\mathbf{a}^T\mathbf{x}, \mathbf{b}^T\mathbf{y}) = \frac{\mathbf{c}^T\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{d}}{\sqrt{\mathbf{c}^T\mathbf{c}}\sqrt{\mathbf{d}^T\mathbf{d}}}. \tag{36}$$

By the Cauchy–Schwarz inequality, we have

$$\begin{aligned}&(\mathbf{c}^T\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2})\mathbf{d}\\ &\le (\mathbf{c}^T\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}\Sigma_{\mathbf{yy}}^{-1/2}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{c})^{1/2}(\mathbf{d}^T\mathbf{d})^{1/2}.\end{aligned} \tag{37}$$

Thus

$$\mathrm{Corr}(\mathbf{a}^T\mathbf{x}, \mathbf{b}^T\mathbf{y}) \le \frac{(\mathbf{c}^T\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{c})^{1/2}}{(\mathbf{c}^T\mathbf{c})^{1/2}} \tag{38}$$

where $\mathbf{c} = \Sigma_{\mathbf{xx}}^{1/2}\mathbf{a}$, and the equality holds if the vectors $\mathbf{d}$ and $\Sigma_{\mathbf{yy}}^{-1/2}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{c}$ are collinear. The maximum correlation is attained if $\mathbf{c}$ is the eigenvector with the maximum eigenvalue of the matrix $\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1/2}$ by Rayleigh quotient [47]. Thus,

$$\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{c}_1 = \lambda_1\mathbf{c}_1$$

$$\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xx}}^{1/2}\mathbf{a}_1 = \lambda_1\Sigma_{\mathbf{xx}}^{1/2}\mathbf{a}_1$$

$$\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\mathbf{a}_1 = \lambda_1\mathbf{a}_1. \tag{39}$$

In other words, $\mathbf{a}_1$ is the eigenvector with the maximum eigenvalue of the matrix $\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}$. The subsequent pairs are found by using eigenvalues of decreasing magnitudes, namely

$$\begin{aligned}&\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}} = Q\Lambda Q^{-1}\\ &= \begin{bmatrix} | & | & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \\ | & | & | \end{bmatrix}\begin{bmatrix}\lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3\end{bmatrix}\begin{bmatrix} - & \mathbf{a}_1^T & - \\ - & \mathbf{a}_2^T & - \\ - & \mathbf{a}_3^T & -\end{bmatrix}\end{aligned} \tag{40}$$

where $|\lambda_1| \ge |\lambda_2| \ge |\lambda_3|$.

Because of the collinearity constraint, we have

$$\mathbf{b}_i \propto \Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\mathbf{a}_i, i \in \{1, 2, 3\}. \tag{41}$$

The canonical correlation coefficients are

$$\rho_i = \mathrm{Corr}(\mathbf{a}_i^T\mathbf{x}, \mathbf{b_i}^T\mathbf{y}), i \in \{1, 2, 3\} \tag{42}$$

which can be rewritten as

$$\begin{aligned}&\mathbf{a}_i^T\Sigma_{\mathbf{xx}}\mathbf{a}_i\mathbf{b}_i^T\Sigma_{\mathbf{yy}}\mathbf{b}_i\rho_i^2 = (\mathbf{a}_i^T\Sigma_{\mathbf{xy}}\mathbf{b}_i)^T\\ &\mathbf{a}_i^T\Sigma_{\mathbf{xx}}\mathbf{a}_i\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\mathbf{a}_i^T\Sigma_{\mathbf{yy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\mathbf{a}_i\rho_i^2\\ &= (\mathbf{a}_i^T\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\mathbf{a}_i)^T\\ &\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\mathbf{a}_i = \rho_i^2\mathbf{a}_i\end{aligned} \tag{43}$$

which means ρ_i^2 is the corresponding eigenvalues

$$\lambda_i = \rho_i^2, i \in \{1, 2, 3\} \tag{44}$$

and

$$\begin{aligned}\sum_{i=1}^{3}\rho_i^2 = \sum_{i=1}^{3}\lambda_i &= \mathrm{Tr}(Q^{-1}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}Q)\\ &= \mathrm{Tr}(\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}Q^{-1}Q)\\ &= \mathrm{Tr}(\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}).\end{aligned} \tag{45}$$

Appendix B
Degenerated Case

When symmetric motion occurs during a period of observation, the motion components projected on two noncolinear direction $\mathbf{l}_1$ and $\mathbf{l}_2$ are colinear, which means

$$\begin{aligned}&\mathbf{l}_1^T(\mathbf{x} - \hat{\mathbf{x}}) \propto \mathbf{l}_2^T(\mathbf{x} - \hat{\mathbf{x}})\\ &\mathbf{l}_1^T(\mathbf{x} - \hat{\mathbf{x}}) = \alpha \cdot \mathbf{l}_2^T(\mathbf{x} - \hat{\mathbf{x}}), \alpha \ne 0\\ &(\mathbf{x} - \hat{\mathbf{x}})^T\mathbf{l}_1 = \alpha \cdot (\mathbf{x} - \hat{\mathbf{x}})^T\mathbf{l}_2, \alpha \ne 0\\ &(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T\mathbf{l}_1 = \alpha \cdot (\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T\mathbf{l}_2, \alpha \ne 0.\end{aligned} \tag{46}$$

Consider that the autocovariance matrix $\Sigma_{\mathbf{xx}} \propto (\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T$, we have

$$\Sigma_{\mathbf{xx}}(\mathbf{l}_1 - \alpha\mathbf{l}_2) = \mathbf{0} \tag{47}$$

since $\mathbf{l}_1$ and $\mathbf{l}_2$ are noncolinear, $\mathbf{l}_1 - \alpha\mathbf{l}_2$ is a nonzero element of the nullspace of $\Sigma_{\mathbf{xx}}$, which means matrix $\Sigma_{\mathbf{xx}}$ is not full

rank, namely, singular. Similarly, $\Sigma_{\mathbf{yy}}$ could also be proven to be singular. Thus, the observability could be guaranteed by checking the determinant of $\Sigma_{\mathbf{xx}}$ and $\Sigma_{\mathbf{yy}}$

$$\det(\Sigma_{\mathbf{xx}}) \neq 0, \det(\Sigma_{\mathbf{yy}}) \neq 0 \tag{48}$$

which is also implictly constrained by the trace correlation calculation [see (8)]. This means that if symmetric motion occurs during a period of observation, the autocovariance matrix $\Sigma_{\mathbf{xx}}$ and $\Sigma_{\mathbf{yy}}$ are both singular.

To further improve the numerical stability against the noisy motion data, we instead check the condition number and the minimum absolute eigenvalue

$$\begin{aligned} \frac{|\lambda_{\max}(\Sigma_{\mathbf{xx}})|}{|\lambda_{\min}(\Sigma_{\mathbf{xx}})|} < \zeta_u, |\lambda_{\min}(\Sigma_{\mathbf{xx}})| > \zeta_b \\ \frac{|\lambda_{\max}(\Sigma_{\mathbf{yy}})|}{|\lambda_{\min}(\Sigma_{\mathbf{yy}})|} < \zeta_u, |\lambda_{\min}(\Sigma_{\mathbf{yy}})| > \zeta_b \end{aligned} \tag{49}$$

where ζ_u is the upper threshold and ζ_b is the bottom threshold, $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ is the maximal and minimal absolute eigenvalue of $\mathbf{A}$. For highly correlated $\mathbf{x}$ and $\mathbf{y}$, the checking condition can be simplied as

$$\frac{|\lambda_{\max}(\Sigma_{\mathbf{xx}})|}{|\lambda_{\min}(\Sigma_{\mathbf{xx}})|} < \zeta_u, |\lambda_{\min}(\Sigma_{\mathbf{xx}})| > \zeta_b \tag{50}$$

or

$$\frac{|\lambda_{\max}(\Sigma_{\mathbf{yy}})|}{|\lambda_{\min}(\Sigma_{\mathbf{yy}})|} < \zeta_u, |\lambda_{\min}(\Sigma_{\mathbf{yy}})| > \zeta_b. \tag{51}$$

REFERENCES

[1] T. Qin, P. Li, and S. Shen, "VINS-Mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018.

[2] H. Cho, Y.-W. Seo, B. V. Kumar, and R. R. Rajkumar, "A multi-sensor fusion system for moving object detection and tracking in urban driving environments," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2014, pp. 1836–1843.

[3] M. Liang, B. Yang, S. Wang, and R. Urtasun, "Deep continuous fusion for multi-sensor 3D object detection," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 641–656.

[4] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *Int. J. Robot. Res.*, vol. 34, no. 3, pp. 314–334, 2015.

[5] S. Shen, N. Michael, and V. Kumar, "Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft MAVs," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2015, pp. 5303–5310.

[6] R. Mur-Artal and J. D. Tardós, "Visual-inertial monocular slam with map reuse," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 796–803, Apr. 2017.

[7] J. Rehder, R. Siegwart, and P. Furgale, "A general approach to spatiotemporal calibration in multisensor systems," *IEEE Trans. Robot.*, vol. 32, no. 2, pp. 383–398, Apr. 2016.

[8] E. Mair, M. Fleps, M. Suppa, and D. Burschka, "Spatio-temporal initialization for IMU to camera registration," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, 2011, pp. 557–564.

[9] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart, "Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2016, pp. 4304–4311.

[10] A. Fertner and A. Sjolund, "Comparison of various time delay estimation methods by computer simulation," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-34, no. 5, pp. 1329–1330, Oct. 1986.

[11] G. Jacovitti and G. Scarano, "Discrete time techniques for time delay estimation," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 525–533, Feb. 1993.

[12] Z. Zhang, Q.-T. Luong, and O. Faugeras, "Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction," *IEEE Trans. Robot. Autom.*, vol. 12, no. 1, pp. 103–113, Feb. 1996.

[13] R. Mur-Artal, J. Montiel, and J. D. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.

[14] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 611–625, Mar. 2018.

[15] J. Zhang and S. Singh, "LOAM: Lidar odometry and mapping in real-time," in *Proc. Robot.: Sci. Syst.*, 2014, vol. 2, p. 9.

[16] C. Kerl, J. Sturm, and D. Cremers, "Dense visual slam for RGB-D cameras," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2013, pp. 2100–2106.

[17] B. Luo and E. Hancock, "Feature matching with procrustes alignment and graph editing," in *Proc. 17th Int. Conf. Image Process. Its Appl.*, 1999, pp. 72–76.

[18] C. Wang and S. Mahadevan, "Manifold alignment using procrustes analysis," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1120–1127.

[19] M. Burri *et al.*, "The Euroc micro aerial vehicle datasets," *Int. J. Robot. Res.*, vol. 35, no. 10, pp. 1157–1163, 2016.

[20] T. D. Larsen, N. A. Andersen, O. Ravn, and N. K. Poulsen, "Incorporation of time delayed measurements in a discrete-time Kalman filter," in *Proc. 37th IEEE Conf. Decis. Control*, 1998, vol. 4, pp. 3972–3977.

[21] E. S. Jones and S. Soatto, "Visual-inertial navigation, mapping and localization: A scalable real-time causal approach," *Int. J. Robot. Res.*, vol. 30, no. 4, pp. 407–430, 2011.

[22] J. Kelly and G. S. Sukhatme, "Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration," *Int. J. Robot. Res.*, vol. 30, no. 1, pp. 56–79, 2011.

[23] S. Weiss, M. W. Achtelik, M. Chli, and R. Siegwart, "Versatile distributed pose estimation and sensor self-calibration for an autonomous MAV," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 31–38.

[24] S. J. Julier and J. K. Uhlmann, "Fusion of time delayed measurements with uncertain time delays," in *Proc. Amer. Control Conf.*, 2005, pp. 4028–4033.

[25] M. Choi, J. Choi, J. Park, and W. K. Chung, "State estimation with delayed measurements considering uncertainty of time delay," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2009, pp. 3987–3992.

[26] F. Zhang and J. Song, "Real-time calibration of gyro-magnetometer misalignment," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 849–856, Apr. 2018.

[27] T. Qin and S. Shen, "Online temporal calibration for monocular visual-inertial systems," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 3662–3669.

[28] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2013, pp. 1280–1286.

[29] M. Fleps, E. Mair, O. Ruepp, M. Suppa, and D. Burschka, "Optimization based IMU camera calibration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2011, pp. 3297–3304.

[30] F. Tungadi and L. Kleeman, "Time synchronisation and calibration of odometry and range sensors for high-speed mobile robot mapping," in *Proc. Australasian Conf. Robot. Autom.*, 2008.

[31] K. Qiu, H. Xie, T. Qin, and S. Shen, "Estimating metric poses of dynamic objects using monocular visual-inertial fusion," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 62–68.

[32] K. Qiu, T. Qin, W. Gao, and S. Shen, "Tracking 3D motion of dynamic objects using monocular visual-inertial sensing," *IEEE Trans. Robotics*, vol. 35, no. 4, pp. 799–816, 2019.

[33] M. Li and A. I. Mourikis, "Online temporal calibration for camera-IMU systems: Theory and algorithms," *Int. J. Robot. Res.*, vol. 33, no. 7, pp. 947–964, 2014.

[34] W. Li and H. Leung, "Simultaneous registration and fusion of multiple dissimilar sensors for cooperative driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 2, pp. 84–98, Jun. 2004.

[35] J. Kelly and G. S. Sukhatme, "A general framework for temporal calibration of multiple proprioceptive and exteroceptive sensors," in *Experimental Robotics*. Berlin, Germany: Springer, 2014, pp. 195–209.

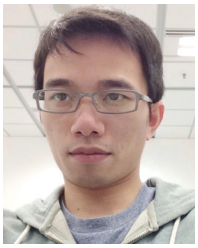
[36] R. Kumar Mishra and Y. Zhang, "A review of optical imagery and airborne Lidar data registration methods," *Open Remote Sens. J.*, vol. 5, no. 1, pp. 54–63, 2012.

[37] L. Zhou, Z. Li, and M. Kaess, "Automatic extrinsic calibration of a camera and a 3D Lidar using line and plane correspondences," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 5562–5569.

[38] D. Scaramuzza, A. Harati, and R. Siegwart, "Extrinsic self calibration of a camera and a 3D laser range finder from natural scenes," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 4164–4169.

[39] J. Marr and J. Kelly, "Unified spatiotemporal calibration of monocular cameras and planar Lidars," in *Proc. IFRR Int. Symp. Exp. Robot.*, Buenos Aires, Argentina, Nov. 5–Dec. 8 2018, pp. 781–790.

[40] F. M. Mirzaei, D. G. Kottas, and S. I. Roumeliotis, "3D Lidar-camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization," *Int. J. Robot. Res.*, vol. 31, no. 4, pp. 452–467, 2012.
[41] G. Pandey, J. R. McBride, S. Savarese, and R. M. Eustice, "Automatic targetless extrinsic calibration of a 3D Lidar and camera by maximizing mutual information," in *Proc. Assoc. Adv. Artif. Intell.*, 2012, pp. 2053–2059.
[42] Z. Taylor and J. Nieto, "Automatic calibration of Lidar and camera images using normalized mutual information," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2013.
[43] J. W. Hooper, "Simultaneous equations and canonical correlation theory," *Econometrica: J. Econometric Soc.*, 1959, pp. 245–256.
[44] B. Thompson, "Canonical correlation analysis," *Encyclopedia of Statistics in Behavioral Science*. Hoboken, NJ, USA: Wiley, 2005.
[45] R. H. Bartels, J. C. Beatty, and B. A. Barsky, *An Introduction to Splines for use in Computer Graphics and Geometric Modeling*. San Mateo, CA, USA: Morgan Kaufmann, 1995.
[46] H. Kim, S. Leutenegger, and A. J. Davison, "Real-time 3D reconstruction and 6-dof tracking with an event camera," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 349–364.
[47] R. A. Horn, R. A. Horn, and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

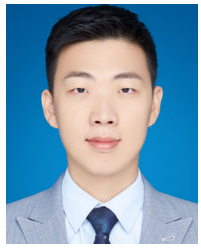


Kejie Qiu (Member, IEEE) received the B.Eng. degree in optical science and engineering from Zhejiang University, Hangzhou, China, in 2013, and the M.Phil. degree in electronic and computer engineering in 2015 from the Hong Kong University of Science and Technology, Hong Kong, where he is currently working toward the Ph.D. degree in electronic and computer engineering.

His research interests include state estimation, 3-D perception, multisensor synchronization, and calibration for autonomous robots.

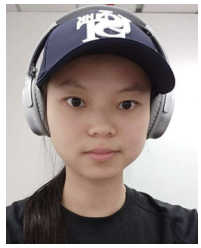
Tong Qin (Graduate Student Member, IEEE) received the B.Eng. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2015. He is currently working toward the Ph.D. degree in electronic and computer engineering with the Hong Kong University of Science and Technology, Hong Kong.

His research interests include state estimation, sensor fusion and visual-inertial localization, and mapping for autonomous robots.



Jie Pan received the B.Eng. degree in software engineering from Xidian University, Xi'an, China, in 2017 and the M.Phil. degree in electronic and computer engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2019.

He is currently a Research Assistant with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology. His research interests include state estimation, sensor fusion, localization and mapping, and autonomous navigation in complex environments.



Siqi Liu received the B.Eng. and M.Eng. degrees in control science and engineering from Harbin Engineering University, Harbin, China, in 2014 and 2017, respectively. She is currently working toward a M.Phil. degree in robotics with the Hong Kong University of Science and Technology, Hong Kong.

Her research interests include state estimation, sensor fusion, object detection, flexible baseline stereo, event based feature tracking.

Shaojie Shen (Member, IEEE) received the B.Eng. degree in electronic engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2009, the M.S. degree in robotics and the Ph.D. degree in electrical and systems engineering, both from the University of Pennsylvania, Philadelphia, PA, USA, in 2011 and 2014, respectively.

He joined the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology in September 2014 as an Assistant Professor. His research interests are in the areas of robotics and unmanned aerial vehicles, with focus on state estimation, sensor fusion, computer vision, localization and mapping, and autonomous navigation in complex environments.