

Non-Parametric Extrinsic and Intrinsic Calibration of Visual-Inertial Sensor Systems

Janosch Nikolic, *Member, IEEE*, Michael Burri, *Student Member, IEEE*, Igor Gilitschenski, *Member, IEEE*, Juan Nieto, *Member, IEEE*, and Roland Siegwart, *Fellow, IEEE*

Abstract—This paper presents a solution for the extrinsic and intrinsic calibration of visual-inertial sensor systems. Calibration is formulated as a joint state and parameter estimation problem of a continuous-time system with discrete-time measurements. A maximum-likelihood estimator is derived to estimate the transform between cameras and inertial sensors, temporal alignment, and inertial sensor intrinsic parameters, such as scale factors, axes misalignment, and sensor noise characteristics. The estimator is simple to implement, consistent, and asymptotically attains the Cramér–Rao lower bound. In contrast to the existing methods, it requires no tuning parameters. Detailed results from repeated calibration experiments with a camera-inertial measurement unit system are reported and compared with the results obtained from a modern, parametric method. We reach a precision of <math><1\text{ mm}</math> in extrinsic translation,

Index Terms—Calibration, inertial measurement units, cameras, maximum likelihood.

I. INTRODUCTION

VISUAL-INERTIAL sensor units are a popular choice for localisation and mapping systems since the combination of these complementary sensing modalities leads to increased robustness and accuracy. However, calibration is required to optimally leverage the measurements of all sensors. Exact knowledge of the transform between the sensors is vital [1]. Most precise systems also align the sensor data temporally to compensate for camera exposure time and inertial measurement unit (IMU) intrinsic delays [2].

While most visual-inertial localisation and mapping frameworks employ an intrinsic calibration for the camera, intrinsic calibration for the inertial sensors is rarely considered in this context (with some notable exceptions, e.g. [3]). To mitigate the problem, factory calibrated IMUs are used [4], or the inertial sensor noise model parameters are tuned to account for unmodelled effects in the gyroscopes and accelerometers [5].

Many modern localisation and mapping systems employ low-cost, micro electromechanical (MEMS) chip gyroscopes

and accelerometers. Due to the technologies used, these can exhibit significant scale factor (sensitivity) and misalignment errors (in the order of 1% and 1° , respectively), see [6]. Once the sensor unit undergoes motion, these terms cause correlated errors which quickly render a purely stochastic sensor model invalid. This leads to errors in the extrinsic calibration and to suboptimal performance.

Our contribution is a probabilistic framework for the joint estimation of sensor extrinsic and IMU intrinsic calibration parameters. Including IMU intrinsic parameters into the entire calibration procedure improves the quality of the calibration, and allows us to achieve an overall performance which is limited largely by the sensor’s noise performances alone. The novel method we present requires no tuning parameters and is optimal with respect to stringent probabilistic criteria. The specification of probabilistic models for those calibration parameters which are considered constant is not necessary.

Instead of exposing the sensors to a precise motion profile, for example on a rate table, calibration is performed using a visual calibration target and arbitrary motion of the sensor unit. Calibration is then treated as a state and parameter estimation problem of a continuous-time, non-linear system with discrete-time, noisy measurements. We develop an estimator based on the maximum likelihood (ML) principle to jointly compute the state trajectory, extrinsic calibration parameters, IMU intrinsic calibration parameters, and sensor time delays. This batch type estimator uses all available measurements and is optimal in the ML sense up to linearisation errors.

Quantitative results from repeated experiments provide an insight into achievable calibration accuracy, and show the typical range of deterministic errors in consumer MEMS gyroscopes and accelerometers. The results also show that the precision and the accuracy of the extrinsic calibration can be improved significantly by incorporating IMU intrinsic calibration terms into the optimisation.

We compare our method directly to more complex, parametric approaches [7]. Parametric approaches use temporal basis functions, for example splines, to represent the motion of the sensor unit and other time varying states such as fluctuating sensor biases. In general, these parametric approaches are regarded as being superior particularly for time delay estimation. However, our experiments reveal that the accuracy of both approaches is comparable. To the best of our knowledge, no such comparison exists in the literature to date.

Manuscript received December 2, 2015; revised March 28, 2016; accepted April 5, 2016. Date of publication April 20, 2016; date of current version June 2, 2016. This work and the development of the visual-inertial sensor unit were supported in part by Armasuisse Science and Technology under Grant 8003501880 and in part by the Commission for Technology and Innovation under Grant 13394.1 PFFLE-NM. The associate editor coordinating the review of this paper and approving it for publication was Dr. Shanhong Xia.

The authors are with the Autonomous Systems Laboratory, ETH Zürich, Zürich CH-8092, Switzerland (e-mail: janosch.nikolic@mavt.ethz.ch; michael.burri@mavt.ethz.ch; igilitschenski@ethz.ch; nietoj@ethz.ch; rsiegwart@ethz.ch).

Digital Object Identifier 10.1109/JSEN.2016.2556662

The remainder of this paper is structured as follows: Related approaches with a focus on works that address IMU intrinsic calibration are summarised in Section II. The sensor models we employ for the calibration of visual-inertial sensor units are described in Section III. The maximum likelihood state and parameter estimator we developed for calibration is presented in Section IV, and the experimental setup and results from repeated calibration experiments are presented in Section V.

II. RELATED WORK

Calibration of camera-IMU systems is receiving considerable attention, mostly due to their use in modern localisation and mapping frameworks. Accordingly, this section presents an overview of existing calibration algorithms with a focus on methods that go beyond the calibration of the relative position and orientation of the sensors. The approaches can be distinguished in terms of i) the type of estimator, e.g., recursive or batch ii) calibration parameters: extrinsics, intrinsics, temporal alignment iii) parametric or non-parametric representations of the sensor unit's motion and bias processes iv) on-line or off-line methods v) calibration infrastructure: natural landmarks, visual calibration target, and other equipment.

Early works by Alves et al. [8] and Lobo and Dias [9] rely on dedicated hardware (a pendulum and a rate table, respectively) to determine the relative pose (position and orientation) between a camera and an IMU, and IMU intrinsic parameters. The different calibration parameters are estimated independently in both approaches, which is suboptimal. In contrast, the method we present requires only a visual calibration target (a checker board, for example), which simplifies the calibration process.

Mirzaei and Roumeliotis [10] present a calibration method based on an extended Kalman filter (EKF), using a visual calibration target and hand-held motion. The calibration parameters (extrinsics) are incorporated into the system state and estimated along with the pose, velocity, and sensor biases. An extension to a full batch solution is also proposed but does not include sensor intrinsic parameters. Kelly and Sukhatme [11] present a similar approach using an unscented Kalman filter. In addition, they propose an extension by using only natural landmarks, without the need for a calibration target. Zachariah and Jansson [12] use a sequential linear filter based on a sigma point Kalman filter and additionally track IMU intrinsic calibration parameters. Li et al. [13] recently proposed an EKF based method to estimate a multitude of calibration parameters of a rolling-shutter camera-IMU setup. Li includes IMU and camera intrinsics, as well as time delays into the system state. This filter based method works on-line and requires no calibration infrastructure.

These methods require the specification of initial uncertainties for the state of the sensor unit as well as the calibration parameters. This is not required for our method. In addition, probabilistic models that describe the evolution of the calibration parameters in time need to be specified and tuned in these works. This is the case since even those calibration parameters which are assumed to be

fixed are part of the state. In contrast, our method allows for parameters to be considered “fixed but unknown” and does not require the specification of a probabilistic model for these parameters.

In contrast to these recursive methods, Hol et al. [14] treat calibration as a gray-box system identification problem. A full state and state covariance of the system are propagated using the IMU, and a linearised cost function with visual residuals (from checker board observations), weighted according to their predicted covariance, is minimised with respect to the camera-IMU extrinsic calibration parameters. This approach is similar to our method, but it does not include IMU intrinsic parameters.

Furgale et al. [2] propose a full maximum likelihood estimator to determine sensor extrinsics and a time delay between camera and IMU. The method uses a parametric representation of the unit's motion and its sensor biases (B-splines). This enables sophisticated calibration features and direct computation of inertial measurement residuals. We will refer to this method as *Kalibr*, since a corresponding toolbox was released under this name.¹ Krebs [15] extends this work by including IMU intrinsic calibration parameters and is similar to our work in terms of sensor modelling. Rehder *et al.* [16] pushes further to leverage the parametric spline representation of the sensor unit's trajectory to determine the position of individual accelerometers. In contrast to [16], we cannot estimate the position of individual accelerometers. We do not estimate angular accelerations which renders our method unsuitable for the estimation of such parameters. However, extensions to our method to address this are feasible.

These parametric methods require that the sensor unit's trajectory and the sensor biases are well approximated with the selected basis functions. They also require the specification of a knot density which in turn depends on the motion. This is not the case for the method presented here. In addition, our method has the advantage that “inertial residuals” are only computed at time instances where visual target points are detected and inertial measurements are integrated in-between.

III. THE VISUAL-INERTIAL SENSOR SYSTEM

A typical visual-inertial sensor system is comprised of one or more cameras C_i which are rigidly connected to a strapdown IMU. We define a moving body coordinate frame B on the sensor unit such that it coincides with the IMU's input reference axes (IRA). Fig. 1 illustrates the sensor system in front of the calibration target T . The corresponding notation is briefly introduced in Section III-A. As the sensor unit moves, the camera observes points from the calibration target through a projection onto its camera image plane, where the observations are corrupted by noise. The corresponding camera measurement model is summarised in Section III-B. At the same time, the noisy gyroscopes and accelerometers measure the sensor unit's angular rate and acceleration. The corresponding gyroscope and accelerometer sensor models are described in Section III-C.

¹*Kalibr* is available at www.github.com/ethz-asl/kalibr (April 2016).

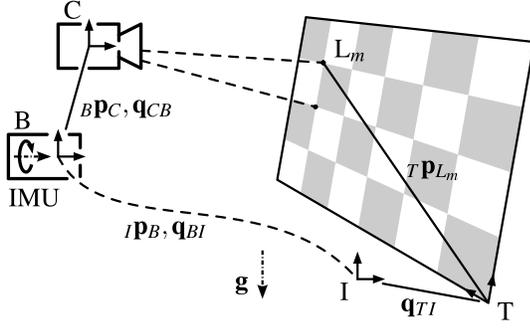


Fig. 1. Visual-inertial sensor system in front of a visual calibration target. The camera C and the IMU are rigidly attached to the sensor unit's body frame B , which coincides with the IMU's input reference axes. Gyroscopes and accelerometers measure the unit's angular rate and acceleration, while the camera observes target points L_m from target T .

A. Notation

We denote a vector \vec{CL} , expressed in B , as ${}_B\mathbf{p}_{CL}$. The vector ${}_B\mathbf{p}_{CL}$ is transformed to an other coordinate frame I as follows

$${}_I\mathbf{p}_{CL} = \mathbf{C}_{IB} {}_B\mathbf{p}_{CL} \quad (1)$$

where \mathbf{C}_{IB} is the direction cosine matrix that transforms vectors from B to I .

We use Hamiltonian unit quaternions \mathbf{q} for a non-minimal but singularity free representation of rotations:

$$\mathbf{q} = [q_w \quad q_x \quad q_y \quad q_z]^T = \begin{bmatrix} q_w \\ \bar{\mathbf{q}} \end{bmatrix} \quad (2)$$

where q_w denotes the real and $\bar{\mathbf{q}}$ the imaginary part of \mathbf{q} . The direction cosine matrix $\mathbf{C}(\mathbf{q})$ is computed from \mathbf{q} as follows

$$\mathbf{C}(\mathbf{q}) = q_w^2 \mathbf{I}_{3 \times 3} + 2q_w [\bar{\mathbf{q}} \times] + [\bar{\mathbf{q}} \times]^2 + \mathbf{q}\mathbf{q}^T \quad (3)$$

where $[\cdot \times]$ denotes the skew-symmetric operator [17]:

$$[\bar{\mathbf{q}} \times] = \begin{bmatrix} 0 & -q_z & q_y \\ q_z & 0 & -q_x \\ -q_y & q_x & 0 \end{bmatrix}. \quad (4)$$

B. Camera Measurement Model

The target points (e.g. checker board corners) ${}_T\mathbf{p}_{TL_m}$, or ${}_T\mathbf{p}_{L_m}$ in short, $m = 1, \dots, M$, are given in the target's reference frame T . The transform between T and the "world" frame I is fixed, but unknown. The translational part is unobservable and hence we set the origin of T to coincide with I . I is aligned with gravity, and the orientation of T with respect to gravity (the nuisance parameter \mathbf{q}_{TI}) is included in the estimation problem to avoid a precise, manual alignment of the calibration target.

For the observation of the m th target point in the unit image plane of the camera at time-step k , $\mathbf{z}_{km} \in \mathbb{R}^2$, we write:

$$\mathbf{z}_{km} = \mathbf{h}(\mathbf{x}_k, \boldsymbol{\theta}) + \mathbf{v} \quad (5a)$$

$$= \boldsymbol{\pi}({}_C\mathbf{p}_{L_m}) + \mathbf{v} \quad (5b)$$

$$= \boldsymbol{\pi}(\mathbf{C}_{CB} (\mathbf{C}_{BI} (\mathbf{C}_{IT} {}_T\mathbf{p}_{L_m} - {}_I\mathbf{p}_B) - {}_B\mathbf{p}_C)) + \mathbf{v} \quad (5c)$$

where $\mathbf{h}(\cdot)$ denotes the camera measurement function, \mathbf{x}_k is the system state at time t_k and includes the position ${}_I\mathbf{p}_B$ and the orientation \mathbf{q}_{BI} of B , and $\boldsymbol{\theta}$ contains all calibration parameters. $\boldsymbol{\pi}(\cdot)$ projects the target point ${}_C\mathbf{p}_{L_m}$, expressed in the camera frame C , to the camera unit image plane, given the camera intrinsic calibration parameters.

The discrete-time, white Gaussian noise process \mathbf{v} is of strength

$$\mathbb{E}[\mathbf{v}_k \mathbf{v}_k^T] = \mathbf{R} \quad (6)$$

with $\mathbf{R} = \sigma_c^2 \mathbf{I}$, where \mathbf{I} is a 2×2 identity matrix. This noise process models uncertainty in the target observations due to sensor noise, motion blur, and discretisation into pixels.

Throughout this article, we assume that the earth-fixed calibration target frame T is not moving with respect to I . Specifically, we assume the earth's rotational rate $\omega_e = 0$. This is a legitimate assumption, given the noise characteristics of the gyroscopes we calibrate. It has the consequence that the orientation of the target is not fully observable, and that only its orientation with respect to the gravity vector can be estimated.

C. Inertial Sensor Model

The gyroscope and accelerometer sensor models have a stochastic and a deterministic component. The stochastic (noise) model we use are described in Section III-C1, and the deterministic models in Section III-C2.

1) *Stochastic Model*: A variety of stochastic processes are used for modelling inertial sensor noise [18]. We employ a simple model that is widely used in the context of visual-inertial sensing (for example in [17], [19], and [20]): a combination of a rapidly fluctuating (white) noise process and a slowly varying, correlated noise process (a bias). A typical source for the wideband noise component in MEMS inertial sensors is electronic noise from transducer and amplifier stages, and fluctuations in drive frequency are a source for bias variation [6].

We denote the gyroscope and accelerometer noise processes as \mathbf{n}_g and \mathbf{n}_a . To keep the notation simple we will not distinguish between random processes and a particular realisation (a sample path) of the process and write

$$\mathbf{n}_g = \mathbf{b}_g + \mathbf{w}_g \quad (7a)$$

$$\mathbf{n}_a = \mathbf{b}_a + \mathbf{w}_a. \quad (7b)$$

\mathbf{w}_g and \mathbf{w}_a denote continuous-time, white Gaussian noise processes of strength σ_g and σ_a ,

$$\mathbb{E}[\mathbf{w}_g(t_1) \mathbf{w}_g(t_2)] = \sigma_g^2 \mathbf{I} \delta(t_1 - t_2) \quad (8a)$$

$$\mathbb{E}[\mathbf{w}_a(t_1) \mathbf{w}_a(t_2)] = \sigma_a^2 \mathbf{I} \delta(t_1 - t_2) \quad (8b)$$

where $\delta(\cdot)$ denotes the Dirac delta function. $\mathbf{b}_g(t)$ and $\mathbf{b}_a(t)$ denote the slowly varying bias processes, with

$$\dot{\mathbf{b}}_g = -\frac{1}{\tau_g} \mathbf{b}_g + \mathbf{w}_{bg} \quad (9a)$$

$$\dot{\mathbf{b}}_a = -\frac{1}{\tau_a} \mathbf{b}_a + \mathbf{w}_{ba} \quad (9b)$$

where \mathbf{w}_{bg} and \mathbf{w}_{ba} are white noise processes of strength σ_{bg} and σ_{ba} , the bias ‘‘diffusions’’.

During calibration we will set the correlation times to $\tau_g = \tau_a = \infty$. We do this for two reasons: i) low-cost MEMS gyroscopes and accelerometers exhibit turn-on biases which remain constant during the operation of the instrument (at constant temperature) but are significant (in the order of a few degrees per second and up to 0.5 ms^{-2}). Setting the correlation times to infinity prevents us from having to estimate the turn-on biases since no a-priori mean is required for an unbounded (i.e. $\tau = \infty$) random walk. ii) an exponentially time-correlated process is rarely a good approximation for true bias variation which often exhibits flickering characteristics. In addition, bias variation is mostly driven by temperature changes which is itself typically not exponentially time-correlated. Nevertheless, exponentially correlated bias processes and turn-on biases, if desired, would fit naturally into the concept presented here.

A viable procedure to obtain the noise strengths (the ‘‘sigmas’’) manually using the Allan variance is given in [18]. We obtain these noise model parameters automatically using the maximum likelihood method outlined in [21]. This reduces the potential for error since it requires no user input. An alternative method is presented in [22].

2) *Deterministic Model*: Various, often identical, deterministic inertial sensor models exist in the literature (see, for example, [15], [23]–[26]). A more generic accelerometer model is proposed in [3], where each accelerometer has an arbitrary orientation and scale on which the specific force is projected.

For tractability, we choose a simple model based on the model of [26]. It incorporates scale errors, axes misalignment (cross axes sensitivity), and gyroscope g-sensitivity:

$$\tilde{\boldsymbol{\omega}}(t - \Delta t_{BC}) = \mathbf{K}_g \mathbf{M}_g {}_B \boldsymbol{\omega}(t) + \mathbf{T}_g \mathbf{a}(t) + \mathbf{n}_g(t) \quad (10a)$$

$$\tilde{\mathbf{a}}(t - \Delta t_{BC}) = \mathbf{K}_a \mathbf{M}_a \mathbf{a}(t) + \mathbf{n}_a(t) \quad (10b)$$

where ${}_B \boldsymbol{\omega}$ denotes the angular rate of the sensor unit B with respect to I , and ${}_B \mathbf{a}$ denotes the specific force, both expressed in B (i.e. the output of ideal gyroscope and accelerometer triads). The scale (\mathbf{K}) and misalignment (\mathbf{M}) matrices (see Fig. 2) are defined as follows:

$$\mathbf{K}_g = \begin{bmatrix} k_{gx} & 0 & 0 \\ 0 & k_{gy} & 0 \\ 0 & 0 & k_{gz} \end{bmatrix} \quad \mathbf{M}_g = \begin{bmatrix} 1 & 0 & 0 \\ \gamma_z & 1 & 0 \\ -\gamma_y & \gamma_x & 1 \end{bmatrix}$$

$$\mathbf{K}_a = \begin{bmatrix} k_{ax} & 0 & 0 \\ 0 & k_{ay} & 0 \\ 0 & 0 & k_{az} \end{bmatrix} \quad \mathbf{M}_a = \begin{bmatrix} 1 & -\alpha_{yz} & \alpha_{zy} \\ \alpha_{xz} & 1 & -\alpha_{zx} \\ -\alpha_{xy} & \alpha_{yx} & 1 \end{bmatrix}.$$

In order to capture a full misalignment between the gyroscope and the accelerometer \mathbf{M}_a contains six small angles rather than only three. This is necessary since none of the gyroscope and accelerometer axes align perfectly in general. \mathbf{T}_g is a fully populated 3×3 matrix that models static gyroscope g-sensitivity.

When data from accurate sources is fused, precise temporal alignment is essential. The visual-inertial sensor unit we used performs a shutter-centric alignment between camera

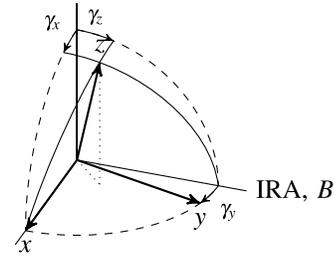


Fig. 2. Misalignment (or cross-axis sensitivity) of the gyroscope triad. The x -axis of the gyroscope is aligned with its corresponding input reference axis (IRA). The (approximate) small angles γ are contained in the misalignment matrix \mathbf{M}_g . Note that the input axes x , y , and z are not orthogonal.

frames and inertial measurements in hardware. IMU intrinsic delays, however, have to be determined from measurements. The main contributor to commercial MEMS gyroscope and accelerometer delays are device internal (digital) low-pass filters. We summarise their effect in one single, constant delay Δt_{BC} . This is an approximation, since the true delay is frequency dependent. Depending on the bandwidth of these filters (often user configurable), typical delays range from below 1 ms to 30 ms or more. This is significant and needs to be taken into account when designing precision visual-inertial estimation algorithms.

IV. CALIBRATION

We treat calibration as a classical state and parameter estimation problem and employ a full batch maximum likelihood (ML) estimator to obtain asymptotically optimal state and parameter estimates (up to linearisation errors).

All unknown static quantities, such as intrinsic and extrinsic calibration parameters, are summarised in $\boldsymbol{\theta}$. Time-varying quantities, such as the pose of the sensor unit or sensor biases, are contained in the state $\mathbf{x}(t)$. The value of $\mathbf{x}(t)$ is estimated at every time instance t_k where a camera frame is acquired. In-between camera frames $\mathbf{x}(t)$ is integrated using the inertial measurements. During integration noise is injected through the gyroscope and accelerometer measurement noise processes.

The states and the parameters which are estimated are summarised in Section IV-A. The equations that link the inertial sensor measurements and the motion of the sensor unit are stated in Section IV-B, and an outline of the state and parameter estimator is given in Section IV-C.

A. System State and Parameters

The estimator we propose jointly estimates the full state trajectory of the system at time instances $k = 0, \dots, K$, $\mathbf{x}(t_k)$, as well as the extrinsic and intrinsic calibration parameters $\boldsymbol{\theta}$.

The system state $\mathbf{x}(t) \in \mathbb{S}^3 \times \mathbb{R}^{12}$ includes the body’s orientation ${}^I \mathbf{q}_{BI}$, its position ${}^I \mathbf{p}_B$, its velocity ${}^I \mathbf{v}_B$, and the gyroscope and accelerometer biases \mathbf{b}_g and \mathbf{b}_a :

$$\mathbf{x} = [{}^I \mathbf{q}_{BI}^T \quad {}^I \mathbf{p}_B^T \quad {}^I \mathbf{v}_B^T \quad \mathbf{b}_g^T \quad \mathbf{b}_a^T]^T. \quad (11)$$

The sensor biases are included in the state since they vary over time in accordance with (9a) and (9b).

TABLE I
THE STATES AND THE PARAMETERS WHICH ARE
ESTIMATED DURING CALIBRATION

	symbol	unit
States (time varying)		
Attitude	\mathbf{q}_{BI}	\mathbb{S}^3 -
Position	${}^I\mathbf{p}_B$	\mathbb{R}^3 m
Velocity	${}^I\mathbf{v}_B$	\mathbb{R}^3 m s^{-1}
Gyroscope bias	\mathbf{b}_g	\mathbb{R}^3 rad s^{-1}
Accelerometer bias	\mathbf{b}_a	\mathbb{R}^3 m s^{-2}
Parameters (fixed)		
Camera-IMU orientation	\mathbf{q}_{CB}	\mathbb{S}^3 -
Camera-IMU translation	${}^B\mathbf{p}_C$	\mathbb{R}^3 m
Camera-IMU time delay	Δt_{BC}	\mathbb{R}^1 s
Target orientation	\mathbf{q}_{TI}	\mathbb{S}^3 -
Gyroscope axes misalignment	$\boldsymbol{\gamma}$	\mathbb{R}^3 rad
Gyroscope scale factors	\mathbf{k}_g	\mathbb{R}^3 -
Gyroscope g-sensitivity	\mathbf{t}_g	\mathbb{R}^9 $\text{rad m}^{-1} \text{s}$
Accelerometer axes misalignment	$\boldsymbol{\alpha}$	\mathbb{R}^6 rad
Accelerometer scale factors	\mathbf{k}_a	\mathbb{R}^3 -

All fixed but unknown extrinsic and intrinsic parameters are collected in $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = [\mathbf{q}_{CB}^T \quad {}^B\mathbf{p}_C^T \quad \Delta t_{BC} \quad \mathbf{q}_{TI}^T \quad \boldsymbol{\gamma}^T \quad \mathbf{k}_g^T \quad \mathbf{t}_g^T \quad \boldsymbol{\alpha}^T \quad \mathbf{k}_a^T]^T.$$

\mathbf{q}_{CB} and ${}^B\mathbf{p}_C$ denote the orientation and the position of the camera with respect to the the body, Δt_{BC} is the time delay between visual and inertial data, and \mathbf{q}_{TI}^T is the orientation of the target with respect to I (i.e. gravity), see Fig. 1. $\boldsymbol{\gamma}$, \mathbf{k}_g , and \mathbf{t}_g contain the gyroscope misalignment small angles, scale factors, and g-sensitivities, and $\boldsymbol{\alpha}$, \mathbf{k}_a are the accelerometer misalignments and scale factors. Table I lists all states and parameters which are estimated during calibration.

The inertial sensor noise model parameters σ_g , σ_{bg} , σ_a , and σ_{ba} are determined automatically prior to calibration and remain fixed (see Section V). The camera target observation uncertainty σ_c is set to the equivalent of $1/5$ pixel to account for sub-pixel accurate target observations.

B. Equations of Motion

The differential equations that govern the motion of the sensor unit can be written as follows

$$\dot{\mathbf{q}}_{BI} = \frac{1}{2} \boldsymbol{\Omega}({}_B\boldsymbol{\omega}) \mathbf{q}_{BI} \quad (12a)$$

$${}^I\dot{\mathbf{p}}_B = {}^I\mathbf{v}_B \quad (12b)$$

$${}^I\dot{\mathbf{v}}_B = \mathbf{C}(\mathbf{q}_{BI})^T {}_B\mathbf{a} + \mathbf{I}\mathbf{g} \quad (12c)$$

where

$$\boldsymbol{\Omega}({}_B\boldsymbol{\omega}) = \begin{bmatrix} 0 & -\omega_x & -\omega_y & -\omega_z \\ \omega_x & 0 & \omega_z & -\omega_y \\ \omega_y & -\omega_z & 0 & \omega_x \\ \omega_z & \omega_y & -\omega_x & 0 \end{bmatrix}.$$

We can substitute the true body angular rates ${}_B\boldsymbol{\omega}$ and accelerations ${}_B\mathbf{a}$ with the noisy, erroneous gyroscope and

accelerometer measurements $\tilde{\boldsymbol{\omega}}$ and $\tilde{\mathbf{a}}$ using the sensor models (10a) and (10b). In summary, we can then write

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}, \mathbf{u}, \mathbf{w}) \quad (13)$$

with $\mathbf{u} = [\tilde{\boldsymbol{\omega}}^T \quad \tilde{\mathbf{a}}^T]^T$, and \mathbf{w} collecting all the driving white noise processes $\mathbf{w} = [\mathbf{w}_g^T \quad \mathbf{w}_{bg}^T \quad \mathbf{w}_a^T \quad \mathbf{w}_{ba}^T]^T$.

C. State and Parameter Estimation

We use the maximum likelihood principle to estimate the state trajectory and the calibration parameters of the sensor unit. In other words we select the state and the parameters such that the probability of observing the measurements that have occurred is maximal. This leads to estimates which are asymptotically unbiased and attain the Cramér-Rao lower bound [27].

Consequently, we aim to solve the following optimisation problem:

$$\hat{\mathbf{X}}_{ML}, \hat{\boldsymbol{\theta}}_{ML} = \underset{\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}}{\operatorname{argmax}} p(\mathbf{Z}, \mathbf{U}, \hat{\mathbf{X}}, \hat{\boldsymbol{\theta}}) \quad (14)$$

where $p(\cdot)$ denotes the joint probability density function. $\hat{\mathbf{X}}_{ML}$ and $\hat{\boldsymbol{\theta}}_{ML}$ denote the maximum likelihood estimates of the state trajectory and the parameters and

$$\hat{\mathbf{X}} = \{\hat{\mathbf{x}}(t_0), \hat{\mathbf{x}}(t_1), \dots, \hat{\mathbf{x}}(t_K)\}$$

$$\mathbf{Z} = \{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_K\}$$

$$\mathbf{U} = \{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_L\}.$$

$\hat{\mathbf{X}}$ and \mathbf{Z} collect the state estimates and camera observations of each visible target point at time instances $k = 0, \dots, K$ (all time instances when a camera frame was captured) and \mathbf{U} contains the sampled inertial sensor measurements at time instances t_0, \dots, t_L .

Since the process noise terms $\mathbf{w} = [\mathbf{w}_g^T \quad \mathbf{w}_{bg}^T \quad \mathbf{w}_a^T \quad \mathbf{w}_{ba}^T]^T$ are white $\mathbf{x}(t)$ is Markov (strictly, this is only true later once we linearised around a nominal state trajectory and nominal parameters, since \mathbf{w} enters the system multiplicatively). Furthermore, since the camera measurement noise process \mathbf{v} is independent, the camera measurements at time instance k only depend on the value of \mathbf{x} at that time t_k . We can therefore factor the joint pdf in (14) as follows:

$$p(\mathbf{Z}, \mathbf{U}, \hat{\mathbf{X}}, \hat{\boldsymbol{\theta}}) = \prod_{k=0}^K \prod_{m=0}^{M-1} p(\mathbf{z}_{km} | \hat{\mathbf{x}}_k, \hat{\boldsymbol{\theta}}) \text{ visual} \\ \times \prod_{k=1}^K p(\hat{\mathbf{x}}_k | \hat{\mathbf{x}}_{k-1}, \mathbf{U}_{k-1}, \hat{\boldsymbol{\theta}}) \text{ inertial} \quad (15)$$

where \mathbf{x}_k denotes the state at time instance k , $\mathbf{x}(t_k)$, \mathbf{z}_{km} denotes the measurement of the m th visual target point at time instance k , and \mathbf{U}_{k-1} collects all gyroscope and accelerometer measurements in the time-interval $[t_{k-1}, t_k]$ (with current estimates of Δt_{BC}); M is the total number of points in the target. $p(\mathbf{z}_{km} | \hat{\mathbf{x}}_k, \hat{\boldsymbol{\theta}})$ is often denoted as the ‘‘measurement model’’ (5a) and $p(\hat{\mathbf{x}}_k | \hat{\mathbf{x}}_{k-1}, \mathbf{U}_{k-1}, \hat{\boldsymbol{\theta}})$ as the ‘‘process model’’ (13).

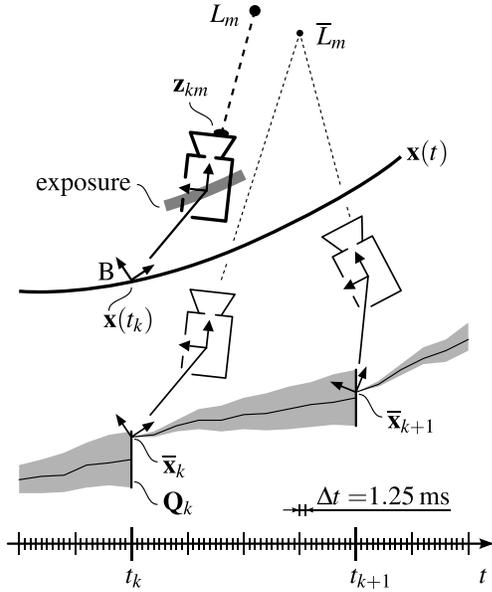


Fig. 3. Illustration of the maximum likelihood state and parameter estimator for calibration. At every time instance t_k a camera frame is captured, providing M target point observations \mathbf{z}_{km} . In-between frames the state is integrated starting from a nominal state trajectory $\bar{\mathbf{x}}_k$ and nominal parameters $\bar{\boldsymbol{\theta}}$, resulting in the visual and inertial error terms in (17a).

1) *Nominal Trajectory and Parameters:* The process and the measurement models are non-linear. We therefore linearise and perturb around a nominal state trajectory $\bar{\mathbf{X}}$ and nominal parameters $\bar{\boldsymbol{\theta}}$ with an “error state” $\delta\mathbf{X}$ and “error parameters” $\delta\boldsymbol{\theta}$:

$$\hat{\mathbf{x}}_k = \bar{\mathbf{x}}_k \boxplus \delta\mathbf{x}_k \quad (16a)$$

$$\hat{\boldsymbol{\theta}} = \bar{\boldsymbol{\theta}} \boxplus \delta\boldsymbol{\theta} \quad (16b)$$

where the \boxplus operator is simply component wise addition, except for the quaternions where quaternion multiplication is used. Fig. 3 shows the sensor system and the relevant quantities.

2) *Minimal State and Parameter Representation:* The attitude of the sensor system B and two calibration parameters, the camera extrinsic orientation and the target orientation, are represented with unit quaternions; a non-minimal representation in \mathbb{S}^3 . We therefore employ minimal coordinates $\delta\boldsymbol{\chi}_k \in \mathbb{R}^{15}$ and $\delta\boldsymbol{\eta} \in \mathbb{R}^{34}$ and mappings $\Psi_{\mathbf{x}}$ and $\Psi_{\boldsymbol{\theta}}$ to transform from minimal coordinates to tangent space [1]:

$$\delta\mathbf{x}_k = \exp(\Psi_{\mathbf{x}}(\delta\boldsymbol{\chi}_k)) \quad (17a)$$

$$\delta\boldsymbol{\theta} = \exp(\Psi_{\boldsymbol{\theta}}(\delta\boldsymbol{\eta})). \quad (17b)$$

The error state dynamics can then be written as:

$$\delta\dot{\boldsymbol{\chi}} = \mathbf{F}(\bar{\mathbf{x}}, \bar{\boldsymbol{\theta}})\delta\boldsymbol{\chi} + \mathbf{G}(\bar{\mathbf{x}})\mathbf{w} \quad (18)$$

where \mathbf{F} denotes the derivative of the state with respect to the minimal coordinates $\delta\boldsymbol{\chi}$ at a particular linearisation point $\bar{\mathbf{x}}, \bar{\boldsymbol{\theta}}$. \mathbf{G} maps the driving white noise processes on the state according to (7a) and (9a).

3) *ML Calibration as a Least Squares Problem:* Instead of maximising the joint pdf (15) directly we minimise its negative logarithm, the “negative log-likelihood function”. Since the system is linearised and the different noise sources

are modelled as independent Gaussians, minimising the log-likelihood becomes a least squares problem:

$$\begin{aligned} \bar{L}(\delta\boldsymbol{\chi}, \delta\boldsymbol{\eta}) = & \sum_{k=0}^K \sum_{m=0}^{M-1} \Delta\mathbf{z}_{km}^T \mathbf{R}^{-1} \Delta\mathbf{z}_{km} \text{ visual} \\ & + \sum_{k=1}^K \Delta\boldsymbol{\chi}_k^T \mathbf{Q}_k^{-1} \Delta\boldsymbol{\chi}_k \text{ inertial} \end{aligned} \quad (19)$$

where $\Delta\mathbf{z}_{km}$ and $\Delta\boldsymbol{\chi}_k$ denote the measurement and (minimal) state residuals. \mathbf{R} is the camera measurement noise covariance (6) and \mathbf{Q}_k denotes the covariance of the state $\delta\boldsymbol{\chi}_k$, given $\delta\boldsymbol{\chi}_{k-1}$ (through the injection of gyroscope and accelerometer noise from time t_{k-1} to t_k).

To linearise we expand around $\bar{\mathbf{X}}$ and $\bar{\boldsymbol{\theta}}$ to first order. For the camera measurements residuals we find

$$\Delta\mathbf{z}_{km} \approx \mathbf{z}_{km} - \left(\mathbf{h}(\bar{\mathbf{x}}_k, \bar{\boldsymbol{\theta}}) + \frac{\partial \bar{\mathbf{h}}}{\partial \delta\boldsymbol{\chi}_k} \delta\boldsymbol{\chi}_k + \frac{\partial \bar{\mathbf{h}}}{\partial \delta\boldsymbol{\eta}} \delta\boldsymbol{\eta} \right) \quad (20a)$$

$$= \Delta\bar{\mathbf{z}}_{km} - \begin{bmatrix} \mathbf{H}_{\boldsymbol{\chi}_k} & \mathbf{H}_{\boldsymbol{\eta}_{km}} \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{\chi}_k \\ \delta\boldsymbol{\eta} \end{bmatrix} \quad (20b)$$

where $\Delta\bar{\mathbf{z}}_{km}$ denotes the measurement residual corresponding to the m th target point at time instance k , $\mathbf{H}_{\boldsymbol{\chi}_k}$ is the measurement Jacobian with respect to the state, and $\mathbf{H}_{\boldsymbol{\eta}_{km}}$ is the derivative of the measurement function for the m th target point at time t_k with respect to the (minimal) calibration parameters (i.e. the camera-IMU extrinsics).

For the state or “inertial” residuals we write

$$\Delta\boldsymbol{\chi}_k \approx \Delta\bar{\boldsymbol{\chi}}_k - \begin{bmatrix} \bar{\boldsymbol{\Phi}}_{k-1} & -\mathbf{I} & \mathbf{J}_{k-1} \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{\chi}_{k-1} \\ \delta\boldsymbol{\chi}_k \\ \delta\boldsymbol{\eta} \end{bmatrix} \quad (21)$$

where $\bar{\boldsymbol{\Phi}}_{k-1}$ denotes the “error state” transition matrix (evaluated at $\bar{\mathbf{X}}, \bar{\boldsymbol{\theta}}$) that takes $\delta\boldsymbol{\chi}$ from t_{k-1} to t_k , given the inertial measurements in this time interval. Different schemes can be used to compute $\bar{\boldsymbol{\Phi}}_{k-1}$ and \mathbf{Q}_k , and we use a first order (Euler) method here. We compute $\bar{\boldsymbol{\Phi}}_{k-1}$ by concatenating state transition matrices computed to first order (i.e. $\Delta t \mathbf{F}_l$) at every IMU measurement instance t_l within t_{k-1} to t_k . The minimal state covariances \mathbf{Q}_k are computed to first order as well. Linear interpolation of the IMU measurements is used to avoid additional time-delay due to the integration. It is also used at the border of the interval. Higher order methods may be used to improve the performance.

The derivative of the state transition with respect to the calibration parameters, \mathbf{J}_{k-1} , would be zero if only the extrinsics or camera intrinsics were estimated. If IMU intrinsic parameters are estimated \mathbf{J}_{k-1} is non-zero. We compute the Jacobians with respect to $\delta\boldsymbol{\eta}$ at each IMU sample time and post multiply them with the remaining state transition matrices up to time t_k to obtain \mathbf{J}_{k-1} . The derivatives with respect to the time delay parameter Δt_{BC} is computed in exactly the same manner.

Finally, we normalise the weighted linear least squares problem (17a) by pre-multiplication with the matrix square root of the inverse of the weighting matrices \mathbf{R} and \mathbf{Q}_k [28].

For the residuals, we write

$$\Delta \mathbf{z}'_{km} = \mathbf{R}^{-\frac{1}{2}} \Delta \mathbf{z}_{km} \quad (22)$$

$$\Delta \boldsymbol{\chi}'_k = \mathbf{Q}_k^{-\frac{1}{2}} \Delta \boldsymbol{\chi}_k \quad (23)$$

and the Jacobians are weighted analogously. The maximum likelihood solution can now be obtained by solving the following over determined, linear least-squares problem:

$$\mathbf{A}' \begin{bmatrix} \delta \boldsymbol{\chi} \\ \delta \boldsymbol{\eta} \end{bmatrix} = \begin{bmatrix} \Delta \boldsymbol{\chi}' \\ \Delta \mathbf{z}' \end{bmatrix} \quad (24)$$

with

$$\mathbf{A}' = \begin{bmatrix} \bar{\boldsymbol{\Phi}}'_0 & -\mathbf{I}'_1 & & & & & \mathbf{J}'_0 \\ & & \ddots & & & & \vdots \\ & & & \bar{\boldsymbol{\Phi}}'_{K-1} & -\mathbf{I}'_K & & \mathbf{J}'_{K-1} \\ \mathbf{H}'_{\chi_0} & & & & & & \mathbf{H}'_{\eta_0} \\ & \mathbf{H}'_{\chi_1} & & & & & \mathbf{H}'_{\eta_1} \\ & & \ddots & & & & \\ & & & & \mathbf{H}'_{\chi_K} & \mathbf{H}'_{\eta_K} \end{bmatrix} \quad (25)$$

where $(\cdot)'$ denotes the normalised (i.e. pre-multiplied with $\mathbf{R}^{-\frac{1}{2}}$ or $\mathbf{Q}_k^{-\frac{1}{2}}$) Jacobians. \mathbf{H}'_{η_k} summarises the normalised camera measurement Jacobians for all target point observations at time instance k .

4) *Algorithm*: To find the maximum likelihood solution for the full non-linear problem we successively linearise, compute the Jacobians, and solve (25) to obtain the least squares solution for $\delta \boldsymbol{\chi}$ and $\delta \boldsymbol{\eta}$. Since (25) is large but very sparse we use a sparse QR solver [29]. The state trajectory and calibration parameter estimates are then updated and the system re-linearised. This procedure is then iterated for a fixed number of iterations (see Section V).

Initial guesses for $\bar{\mathbf{X}}$ and $\bar{\boldsymbol{\theta}}$ must be provided to bootstrap the estimator. In general the calibration parameters can be initialised with the default values, i.e. unit scale factors, no misalignment, and no g-sensitivity, and coarse extrinsics (no translation and time shift but an approximate orientation between camera and IMU). We initialise the orientation and position in $\bar{\mathbf{X}}$ with pose estimates obtained from target observations alone (where target observations are available and “zero order hold” in-between), in conjunction with the initial guesses for the camera-IMU extrinsics. Body velocities and sensor biases are initialised to zero. Step-size control or robust cost functions were not necessary for the calibration problems we encountered.

V. EXPERIMENTS AND RESULTS

Calibration experiments were conducted to validate our framework and assess its performance. Ground truth was obtained where possible to evaluate the accuracy of our method. Ground truth could not be obtained for the IMU intrinsic calibration parameters. Where ground truth was

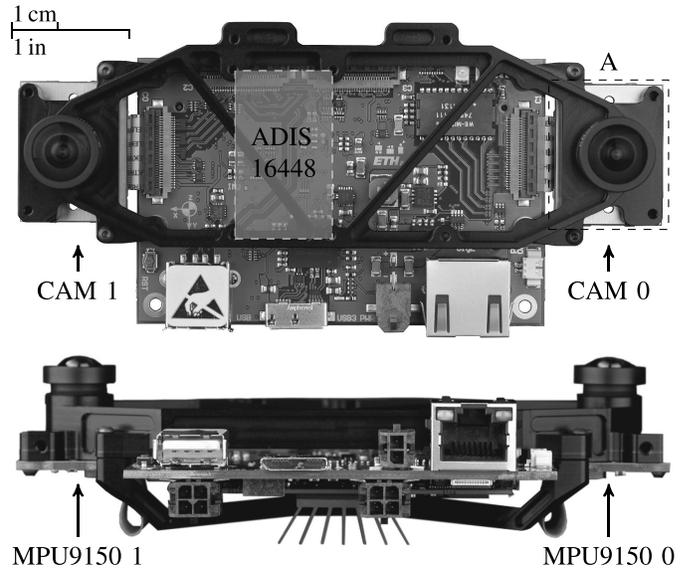


Fig. 4. Visual-inertial sensor unit [4] used for assessing the accuracy and precision of the proposed calibration method. An MPU9150 MEMS IMU is located behind each of the cameras, and a factory calibrated ADIS16448 IMU was attached to the unit. Photo: François Pomerleau.

not obtainable repeated trials were conducted to assess the *precision* of the estimator. Biases introduced through modelling errors are in these cases not assessed. In addition, the results were compared quantitatively with a competing, state of the art parametric calibration method (an extended version of Kalibr, denoted here as *EKalibr* [16]). We refer to our method as discrete maximum likelihood (DML) here.

Section V-A describes the experimental setup that was used for the experiments. Section V-B reports the inertial sensor noise model parameters that were used in the estimator and how they were determined. Section V-C reports calibration results for the camera-IMU extrinsics, Section V-D analyses the temporal alignment, and Section V-E the IMU intrinsic parameter estimates.

A. Experimental Setup

We now describe the setup that was used to conduct the experiments. It is important to highlight that temperature variation has an important effect on most inertial sensors [30]. However, this is outside the scope of this report; calibration is performed at a constant temperature only.

1) *Hardware Setup*: We used a visual-inertial sensor unit that provided hardware synchronised measurements from two cameras and three IMUs. Fig. 4 depicts the unit. A factory calibrated ADIS16448 MEMS IMU from Analog Devices [31] and the two consumer grade MPU9150 chip MEMS IMUs from Invensense [32] provided gyroscope and accelerometer measurements at a rate of 800 Hz each. The cameras consisted of MT9V034 global shutter CMOS chips from Aptina in conjunction with BM2820 S-mount lenses with a diagonal field of view of 122°. The cameras were intrinsically calibrated prior to data collection using the camera intrinsic calibration functionality of Kalibr. Images were captured at a rate of 20 frames per second per camera. The cameras were pre-triggered in order to temporally align all sensor data to

the center of the camera exposure time – a core feature of the sensor unit’s FPGA firmware developed by Nikolic *et al.* [4].

2) *Data Collection*: The sensor unit was moved in front of a checkerboard with 6×7 corners and a corner spacing of 60 mm for dataset collection. A total of 50 datasets with a length 20 s each were captured in sequence and stored for post-processing.

3) *Estimator*: The camera noise parameter σ_c was set to the equivalent of $1/5$ pixel to account for the sub-pixel accurate target point detections. The number of iterations in the algorithm was fixed to 20, step-size control and robust residual weighting were disabled. Spurious target point detections were, however, discarded prior to calibration based on a consistency check using camera observations alone.

4) *Kalibr*: Kalibr was run in its extended version [16]. It was configured to use the same measurements and the exact same noise parameter settings that were used in the estimator presented here.

B. Noise Model Parameters

The estimator we present requires gyroscope and accelerometer noise model parameters in order to compute \mathbf{Q}_k and ultimately a valid maximum likelihood estimate of the calibration parameters. If the sensor noise models are incorrect the estimator will be inconsistent, the results suboptimal, and derived parameter covariances incorrect.

It is often difficult to obtain these parameters unambiguously from data sheets particularly for low-cost devices. We therefore used the method presented in [21] to automatically determine the IMU noise model parameters for all axes of each sensor. The models were identified from sensor data captured when the sensor unit was at rest at constant room temperature. An alternative procedure to obtain these parameters using the Allan variance is presented in [18]. This procedure can lead to similar results but it is a manual method.

We picked the noise model parameters that corresponded to the worst performing sensor axis per gyroscope and accelerometer triad and then used these parameters for all axes. This is not strictly necessary and selecting individual noise model parameters instead would not increase the computational complexity. This is the case since \mathbf{Q}_k in (23) cannot be pre-computed in general. Fig. 5 shows the sample Allan deviation (grey) for the gyroscopes and accelerometers of the MPU9150 (MPU 0). The synthetic Allan deviations corresponding to the model parameters used by the estimator are shown in black (solid). The model’s expressive power is not large enough to capture the long-term bias fluctuations accurately but captures the noise behaviour well for short correlation times. Table II lists the identified noise model parameters used for the gyroscopes and accelerometers by the estimator. Note that the MPU9150, a $4 \times 4 \times 1$ mm consumer IMU, compares favourably with the ADIS16448 in terms of noise performance.

C. Extrinsic

The translation and the rotation between camera and IMU are key parameters in visual-inertial sensor systems.

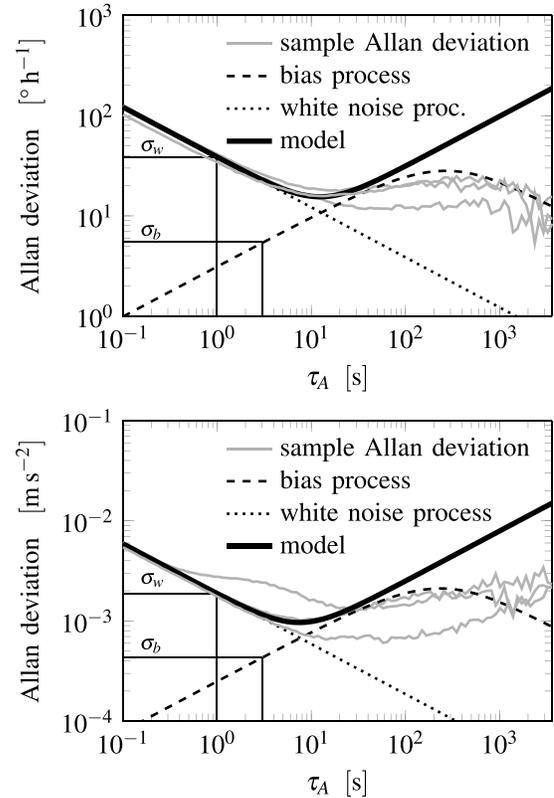


Fig. 5. Allan deviation of the gyroscopes (top) and accelerometers (bottom) of the MPU9150 MEMS IMU. Conventional sample Allan deviations [33] for each sensor axis are shown in grey and Allan deviations corresponding to the noise models used for the experiments are shown in black.

TABLE II
IDENTIFIED GYROSCOPE AND ACCELEROMETER NOISE MODEL
PARAMETERS FOR THE ADIS16448 AND THE MPU9150
MEMS IMUS AS USED FOR THE EXPERIMENTS

		ADIS16448	MPU9150	unit
Gyroscopes				
White noise str.	σ_g	1.86×10^{-4}	8.94×10^{-5}	$\text{rad s}^{-1} \sqrt{\text{Hz}}^{-1}$
Bias diffusion	σ_{bg}	2.66×10^{-5}	1.08×10^{-5}	$\text{rad s}^{-2} \sqrt{\text{Hz}}^{-1}$
Bias corr. time	τ_b	∞	∞	s
Accelerometers				
White noise str.	σ_a	1.86×10^{-3}	2.24×10^{-3}	$\text{m s}^{-2} \sqrt{\text{Hz}}^{-1}$
Bias diffusion	σ_{ba}	4.33×10^{-4}	7.53×10^{-5}	$\text{m s}^{-3} \sqrt{\text{Hz}}^{-1}$
Bias corr. time	τ_b	∞	∞	s

The estimates for the corresponding calibration parameters are shown in Table III. The column “intr. off” refers to the estimates when sensor intrinsic estimation is disabled (but not the delay, which is always estimated), and “intr. on” to when they were enabled.

The results for the MPU 0 and MPU 1 show that even for short calibration sequences in the order of 20s, a precision of <1 mm on all axes in translation and <1 mrad in rotation can be achieved if IMU intrinsics are estimated (see column “intr. on” in Table III). The precision for the ADIS is lower which is most likely due to i) its poorer noise performance

TABLE III

RESULTS FOR THE EXTRINSIC CALIBRATION PARAMETER ESTIMATES FOR THE MPU9150 0, MPU9150 1, AND THE ADIS16448, WITH RESPECT TO CAM 0. THE STANDARD DEVIATIONS ($\pm 1\sigma$) ARE COMPUTED OVER 50 CALIBRATION DATASETS AND GIVE AN INDICATION OF THE PRECISION OF THE ESTIMATOR, BUT NOT ITS ACCURACY

device	parameter		intr. off	intr. on
MPU 0	β_{PC} [mm]	x	0.60 ± 1.46	-5.15 ± 0.58
		y	4.41 ± 1.19	7.92 ± 0.63
		z	-19.33 ± 1.25	-19.54 ± 0.65
	β_{CB} [mrad]	β_x	4.80 ± 1.26	6.78 ± 0.61
		β_y	-0.06 ± 1.16	5.34 ± 0.54
		β_z	5.52 ± 0.68	3.86 ± 0.51
MPU 1	β_{PC} [mm]	x	0.63 ± 1.49	-3.73 ± 0.55
		y	118.25 ± 1.06	113.04 ± 0.60
		z	-26.78 ± 0.71	-18.60 ± 0.62
	β_{CB} [mrad]	β_x	-1.45 ± 1.02	-7.65 ± 0.52
		β_y	0.51 ± 1.37	5.23 ± 0.43
		β_z	6.70 ± 0.37	7.43 ± 0.53
ADIS	β_{PC} [mm]	x	34.02 ± 1.84	34.58 ± 1.91
		y	-11.95 ± 1.35	-9.05 ± 2.86
		z	0.95 ± 0.92	1.69 ± 3.52
	γ_{CB} [mrad]	γ_x	0.95 ± 1.16	1.42 ± 2.45
		γ_y	2.98 ± 1.27	4.42 ± 1.86
		γ_z	7.37 ± 0.53	5.77 ± 1.62

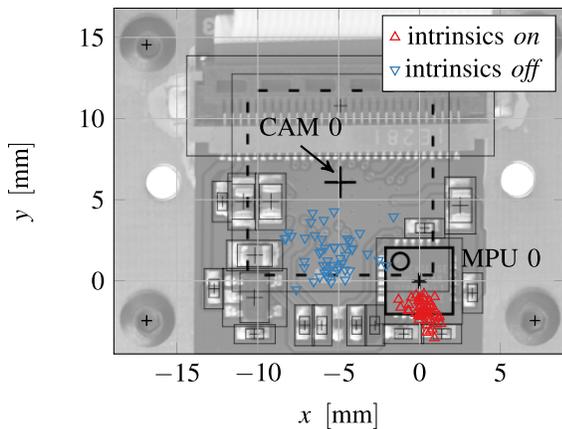


Fig. 6. Bottom view of the camera-IMU printed circuit board of the visual-inertial sensor unit (section “A” in Fig. 4). Translation extrinsics estimates, expressed in the camera reference frame, are indicated for calibration with (Δ) and without (∇) IMU intrinsics.

and ii) the fact that the ADIS relies on separate accelerometers which results in “size effects” [34] (which we do not model).

Fig. 6 depicts the printed circuit board (PCB) of the visual-inertial sensor units camera-IMU module (section “A” in Fig. 4) with the MPU9150 MEMS IMU and the Aptina MT9V034 CMOS camera chip (on top layer, not visible). Extrinsic translation estimates in x and y from all 50 calibration experiments, expressed in the camera reference frame, are indicated for calibration with (Δ) and without (∇)

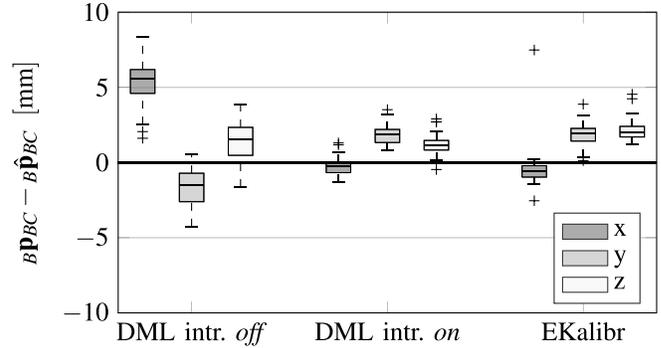


Fig. 7. Camera-IMU extrinsic translation estimation error statistics using ground truth from CAD data to assess the accuracy of the calibration. The error statistics correspond to the translation between CAM 0 and MPU9150 0 and refer to the center of the MPU’s sensor package.

TABLE IV

MEAN AND STANDARD DEVIATION OF THE TIME DELAY ESTIMATES $\Delta \hat{t}_{BC}$ OF THE MPUs WITH RESPECT TO CAM 0, OBTAINED BY THE METHOD PROPOSED HERE (DML) AND *EKalibr*

Method	MPU9150 0	MPU9150 1
DML intrinsics off	$3.0226 \text{ ms} \pm 23.96 \mu\text{s}$	$2.9886 \text{ ms} \pm 15.17 \mu\text{s}$
DML intrinsics on	$3.0306 \text{ ms} \pm 9.59 \mu\text{s}$	$3.0486 \text{ ms} \pm 8.38 \mu\text{s}$
EKalibr	$3.0305 \text{ ms} \pm 15.45 \mu\text{s}$	$3.0321 \text{ ms} \pm 8.27 \mu\text{s}$

IMU intrinsics. These results indicate that the extrinsics are estimated more accurately when the IMU intrinsics are taken into consideration, removing a significant bias and improving the precision of the estimates significantly. We assume that the accelerometers of the MPU9150 are not located in the center of the sensor package which could explain the offset of 1.87 mm (in mean) along the y axis.

Fig. 7 summarises the calibration results for the extrinsic translation between CAM 0 and MPU 0. The corresponding ground truth was obtained from CAD data and is given with respect to the center of the IMU’s package. The figure includes calibration results from *EKalibr* which is also capable of estimating IMU intrinsic calibration parameters.

Despite the different nature of the two algorithms the accuracy and the precision of both methods is comparable. The estimates obtained with *EKalibr* contained a small number of outliers not all of which are visible in Fig. 7.

D. Time Delay

Table IV summarises the time delay estimates we obtained using the method presented here (DML) with and without IMU intrinsics. It also shows the results obtained with *EKalibr*. Outliers from the *EKalibr* estimates were removed prior to computing the standard deviations across all calibration experiments.

The precision of the time delay estimates increases when IMU intrinsic parameters are incorporated into the calibration problem (“intrinsics on”). It is in this case in the order of $10 \mu\text{s}$ which corresponds to less than one-hundredth of the IMU sampling time of 1.25 ms. This is surprising given i) the different nature of the estimators, and ii) that different

TABLE V
IMU INTRINSIC CALIBRATION PARAMETER ESTIMATES
(DML INTRINSICS ON) FOR THE MPU9150 0, THE MPU9150 1,
AND THE ADIS16448. INDICATED ARE THE MEAN AND
THE STANDARD DEVIATION OVER 50 CALIBRATION
EXPERIMENTS USING THE PROPOSED METHOD

param.		ADIS	MPU0	MPU1
Gyroscopes				
$\Delta \mathbf{k}_g$ [%]	k_x	0.68 ± 2.71	2.18 ± 0.59	-7.45 ± 0.56
	k_y	-0.54 ± 2.95	-9.77 ± 0.51	-2.11 ± 0.59
	k_z	2.28 ± 1.16	2.33 ± 0.25	-3.00 ± 0.24
\mathbf{M}_g [mrad]	γ_x	7.83 ± 3.19	1.36 ± 0.42	-1.78 ± 0.44
	γ_y	0.73 ± 2.11	8.81 ± 0.59	9.54 ± 0.47
	γ_z	1.28 ± 3.01	-16.30 ± 0.47	24.64 ± 0.42
Accelerometers				
$\Delta \mathbf{k}_a$ [%]	k_x	21.60 ± 6.46	-1.39 ± 1.51	-5.93 ± 1.93
	k_y	-8.45 ± 13.72	6.19 ± 1.03	1.39 ± 0.89
	k_z	-3.41 ± 7.03	5.74 ± 1.58	6.37 ± 1.80
\mathbf{M}_a ["]	α_{xz}	3.12 ± 6.42	0.55 ± 2.39	-0.31 ± 2.41
	α_{yx}	1.51 ± 4.47	-0.33 ± 2.17	5.73 ± 2.61
	α_{yx}	40.20 ± 16.73	-2.38 ± 1.38	-2.02 ± 1.25
	α_{yz}	-7.92 ± 15.19	-7.29 ± 1.16	5.07 ± 1.28
	α_{zy}	-0.79 ± 7.65	-5.31 ± 1.01	-4.39 ± 1.37
	α_{zx}	-10.44 ± 7.51	9.57 ± 1.39	-15.63 ± 1.03

effects are lumped into the time delay parameter. It also highlights the importance of a precise temporal alignment between the camera and the gyroscopes. This is particularly true for systems that exhibit fast dynamics or are subject to vibrations.

The time delay estimates from DML (intrinsic on) and *EKalibr* are almost identical and have a similar precision. Therefore, both approaches are suitable for precise temporal alignment.

E. IMU Intrinsic

The estimates of the IMU intrinsic calibration parameters are shown in Table V. The misalignment and scale factor errors of the MPU's gyroscopes are in the order of up to 1% and 1.5°, respectively, which is significant. The misalignment and scale factor errors of the MPU's accelerometers appear to be slightly lower. The estimation results for the g-sensitivities of the gyroscopes (not shown) are less conclusive.

The differences between the MPU 0 and the MPU 1 intrinsic parameters highlight that the intrinsic calibration parameters are device specific. The data sheet of the MPU9150 specifies a scale factor tolerance of $\pm 3\%$ and a cross-axis sensitivity of $\pm 2\%$ for the gyroscopes. The parameters we estimated are within this specification.

VI. CONCLUSION AND FUTURE WORK

We presented a novel estimator for extrinsic and intrinsic calibration of visual-inertial sensor systems. The method is accurate and requires no tuning parameters; it works directly with realistic, automatically obtained sensor noise models. We showed that the quality of the calibration improves significantly once IMU intrinsic parameters are included in the estimator. The magnitude of the estimated intrinsic

calibration parameters, including time-delay, highlighted the importance of an IMU intrinsic calibration.

Including angular velocity and angular acceleration in the system state would be a useful extension of this work and facilitate estimation in more general inertial sensor configurations. Furthermore, IMU intrinsic calibration parameters are temperature dependent and not stable over the life time of the sensor unit. Thus an extension to automatic in-filed calibration will be considered in future work.

ACKNOWLEDGMENT

First and foremost, the authors wish to express their gratitude to Joern REHDER of the Autonomous Systems Lab, ETH Zürich, for scientific guidance and support with all aspects of this work. Dr. Paul Furgale, Amir Melzer, and Christian Krebs have also had a profound effect on the theoretical part of this work. The engineering efforts of Pascal Gohl are gratefully acknowledged, and we thank all the *Kalibr* contributors.

REFERENCES

- [1] 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.
- [2] 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. (IROS)*, Nov. 2013, pp. 1280–1286.
- [3] M. Li and A. I. Mourikis, "Online temporal calibration for camera-IMU systems: Theory and algorithms," *Int. J. Robot. Res.*, vol. 33, pp. 947–964, May 2014.
- [4] J. Nikolic *et al.*, "A synchronized visual-inertial sensor system with FPGA pre-processing for accurate real-time SLAM," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May/June 2014, pp. 431–437.
- [5] S. Weiss, D. Scaramuzza, and R. Siegwart, "Monocular-SLAM-based navigation for autonomous micro helicopters in GPS-denied environments," *J. Field Robot.*, vol. 28, no. 6, pp. 854–874, Nov. 2011.
- [6] V. Kempe, *Inertial MEMS: Principles and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [7] P. Furgale, T. Barfoot, and G. Sibley, "Continuous-time batch estimation using temporal basis functions," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Saint Paul, MN, USA, May 2012, pp. 2088–2095.
- [8] J. Alves, J. Lobo, and J. Dias, "Camera-inertial sensor modeling and alignment for visual navigation," in *Proc. 11th Int. Conf. Adv. Robot.*, 2003, pp. 1693–1698.
- [9] J. Lobo and J. Dias, "Relative pose calibration between visual and inertial sensors," *Int. J. Robot. Res.*, vol. 26, no. 6, pp. 561–575, 2007.
- [10] F. M. Mirzaei and S. I. Roumeliotis, "A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation," *IEEE Trans. Robot.*, vol. 24, no. 5, pp. 1143–1156, Oct. 2008.
- [11] 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.
- [12] D. Zachariah and M. Jansson, "Joint calibration of an inertial measurement unit and coordinate transformation parameters using a monocular camera," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2010, pp. 1–7.
- [13] M. Li, H. Yu, X. Zheng, and A. I. Mourikis, "High-fidelity sensor modeling and self-calibration in vision-aided inertial navigation," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May/June 2014, pp. 409–416.
- [14] J. D. Hol, T. B. Schön, and F. Gustafsson, "Modeling and calibration of inertial and vision sensors," *Int. J. Robot. Res.*, vol. 29, nos. 2–3, pp. 231–244, 2010.
- [15] C. Krebs. (2012). *Generic IMU-Camera Calibration Algorithm*. [Online]. Available: <http://www.asl.ethz.ch/>.
- [16] J. Rehder, J. Nikolic, T. Hinzmann, and R. Siegwart, "Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016.

- [17] N. Trawny and S. I. Roumeliotis, "Indirect Kalman filter for 3D attitude estimation," Dept. Comput. Sci. Eng., Univ. Minnesota, Minneapolis, MN, USA, Tech. Rep. 2005-002, 2005.
- [18] *IEEE Standard Specification Format Guide and Test Procedure for Single-Axis Interferometric Fiber Optic Gyros*, IEEE Standard 952-1997, IEEE Aerospace and Electronic Systems Society, 1998.
- [19] S. Leutenegger, P. Furgale, V. Rabaud, M. Chli, K. Konolige, and R. Siegwart, "Keyframe-based visual-inertial SLAM using nonlinear optimization," in *Proc. Robot., Sci. Syst. (RSS)*, Berlin, Germany, 2013.
- [20] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Consistency analysis and improvement of vision-aided inertial navigation," *IEEE Trans. Robot.*, vol. 30, no. 1, pp. 158–176, Feb. 2014.
- [21] J. Nikolic, P. Furgale, A. Melzer, and R. Siegwart, "Maximum likelihood identification of inertial sensor noise model parameters," *IEEE Sensors J.*, vol. 16, no. 1, pp. 163–176, Jan. 2016.
- [22] Y. Stebler, S. Guerrier, J. Skaloud, and M.-P. Victoria-Feser, "Generalized method of wavelet moments for inertial navigation filter design," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 2, pp. 2269–2283, Jul. 2014.
- [23] A. Chatfield, *Fundamentals of High Accuracy Inertial Navigation*. Washington, DC, USA: AIAA, 1997.
- [24] R. M. Rogers, *Applied Mathematics in Integrated Navigation Systems*. Washington, DC, USA: AIAA, 2007.
- [25] D. H. Titterton and J. L. Weston, *Strapdown Inertial Navigation Technology*. London, U.K.: IEE, 2004.
- [26] I. Skog and P. Händel, "Calibration of a MEMS inertial measurement unit," in *Proc. 17th IMEKO World Congr.*, Rio de Janeiro, Brazil, 2006, pp. 1–6.
- [27] P. S. Maybeck, *Stochastic Models, Estimation, and Control*. New York, NY, USA: Academic, 1979.
- [28] F. Dellaert and M. Kaess, "Square root SAM: Simultaneous localization and mapping via square root information smoothing," *Int. J. Robot. Res.*, vol. 25, no. 12, pp. 1181–1203, 2006.
- [29] T. A. Davis, *Direct Methods for Sparse Linear Systems*. London, U.K.: SIAM, 2006.
- [30] Y. Yuksel, N. El-Sheimy, and A. Noureldin, "Error modeling and characterization of environmental effects for low cost inertial MEMS units," in *Proc. IEEE/ION Position Location Navigat. Symp. (PLANS)*, May 2010, pp. 598–612.
- [31] *Compact, Precision Ten Degrees of Freedom Inertial Sensor ADIS16448*, Analog Devices, Cambridge, MA, USA, 2014.
- [32] *MPU9150 MEMS IMU Product Specification Revision 4.3*, InvenSense, San Jose, CA, USA, 2013.
- [33] D. W. Allan, "Statistics of atomic frequency standards," *Proc. IEEE*, vol. 54, no. 2, pp. 221–230, Feb. 1966.
- [34] J. C. Hung, J. S. Hunter, W. W. Stripling, and H. V. White, "Size effect on navigation using a strapdown IMU," Technol. Lab., U.S. Army Missile Res. Develop. Command, Huntsville, AL, USA, Tech. Rep. T-79-73, 1979.



Janosch Nikolic received the diploma (Dipl. El. Ing. FH) in electrical engineering from the HSR, Rapperswil, Switzerland, and the M.Sc. degree in signal processing and communications from the University of Edinburgh, Edinburgh, U.K. He is currently working towards the Ph.D. degree under the supervision of Prof. Roland Siegwart at the Autonomous Systems Laboratory, ETH Zürich, Zürich, Switzerland. His research is focused on the design, characterization, and calibration of visual-inertial sensor systems with application to localization and mapping, and robot guidance.



Michael Burri received the M.Sc. degree in robotics and control from the Swiss Federal Institute of Technology, Zürich, in 2011. He is currently working towards the Ph.D. degree at the Autonomous Systems Laboratory, ETH Zürich, Zürich, Switzerland, and is a member of the laboratory's rotary wing research group. His research interests are in control, state estimation, system identification, and planning for microaerial vehicles, with the goal of fully automated flight using IMU and on-board cameras as the main sensors.



domains.

Igor Gilitschenski received the Ph.D. degree from the Intelligent Sensor-Actuator-Systems Laboratory, Karlsruhe Institute of Technology (KIT), Germany, in 2015. Before joining the KIT, he obtained his diploma degree in mathematics from the University of Stuttgart, Germany. He is currently a Post-Doctoral Researcher with the Autonomous Systems Laboratory, ETH Zürich, Zürich, Switzerland. His research interests include robotic perception and dynamic state estimation, with a focus on nonlinear systems and nonlinear



international journals and conferences. He is currently an Associate Editor of the IEEE ROBOTICS AND AUTOMATION LETTERS and has previously served as an Editor of the IEEE International Conference on Robotics and Automation, the IEEE International Conference on Intelligent Robots and Systems, and the Robotics Science and Systems, among others.

Juan Nieto received the B.Sc. degree in electronics engineering from Universidad Nacional del Sur, Argentina, in 2000, and the Ph.D. degree in Robotics from the University of Sydney in 2005. He is the Deputy Director at the Autonomous Systems Laboratory, ETH Zürich, Zürich, Switzerland. Before joining ETH, he was a Senior Research Fellow at the Australian Centre for Field Robotics. His main research interest is in navigation, perception, data fusion, and machine learning for robotics. He has over 70 scientific publications in



of around 30 people working on several aspects of robotics. Prof. Siegwart is a member of the Swiss Academy of Engineering Sciences and the Officer of the International Federation of Robotics Research. He has served as the Vice President for Technical Activities from 2004 to 2005, a Distinguished Lecturer from 2006 to 2007, and an AdCom member from 2007 to 2009 of the IEEE Robotics and Automation Society.

Roland Siegwart received the M.Sc. and Ph.D. degrees in mechanical engineering from ETH Zürich, Zürich, Switzerland. He has been a Full Professor for Autonomous Systems with ETH Zürich, Zürich, Switzerland, since 2006, and the Vice President of Research and Corporate Relations since 2010. From 1996 to 2006, he was an Associate Professor and later a Full Professor for Autonomous Microsystems and Robotics with the Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland. He leads a research group