This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2017.2674307, IEEE Sensors Journal

Camera/IMU Calibration Revisited

Joern Rehder and Roland Siegwart[1]

***Abstract*—With growing interest in visual/inertial state estimation and an increasing number of approaches and applications emerging for this technology, camera/IMU calibration can be a valuable tool to increase the performance of these methods and to further the understanding of the involved sensor modalities.**

In this work, we assess the impact of two different adjustments to the commonly used sensor models: First, we extend the IMU model to take the displacement of individual accelerometer axes into account. We show that especially high quality devices benefit from this extension, since these IMUs often employ separate sensors for each axis. Second, we propose a novel, direct model for the camera measurements that operates on image intensities rather than corner positions. This formulation is capable of explicitly accounting for motion blur and defocus, but it requires significant modelling efforts.

Our results demonstrate that the transformation between camera and IMU can be estimated to a precision exceeding $\frac{1}{5}$mm and $\frac{1}{100}°$, while temporal offsets are determined to microsecond precision—on datasets of merely 20 second length. At the same time, image exposure time can be inferred to an accuracy of about $\frac{2}{100}$ ms from motion blur.

I. Introduction

With an increasing number of approaches emerging which leverage the complementary strengths of inertial measurement units (IMUs) and cameras (e. g. [1]–[3]), camera/IMU extrinsic calibration has equally seen a surge in interest. Among the different approaches, offline methods that estimate the relative orientation and displacement between camera and IMU from data collected while moving the sensor suite in front of a stationary calibration target have gained most traction in the robotics community ([4]–[6]). Early on, temporal offsets have been identified as a significant source of deterministic error ([7], [8]). Consequently, the estimation of temporal quantities has been incorporated as an integral part into camera/IMU calibration ([9], [10]).

Similarly, incorrect IMU intrinsics have been eyed as a factor that limits calibration precision, with a number of approaches extending calibration to include more comprehensive inertial measurement models ([10]–[12]).

Recently, different approaches for online calibration of camera/IMU systems have been proposed. Some methods are limited to the transformation between camera and IMU ([2], [3]) while others additionally also determine the time offset [13] and IMU and camera intrinsics [14]. Online approaches exhibit distinct advantages for volatile parameters. In contrast, offline approaches benefit from controlled environments with dedicated calibration motion and are able to expend significantly more computation, which enables batch solutions over large sets of measurements. For these reasons, offline approaches can potentially yield more accurate results for constant parameters.

This work revisits the topic of offline camera/IMU calibration for a more in-depth view at sensor modelling.

With respect to the IMU model, we show that the displacement of individual accelerometers, sometimes referred to as *size-effect* [15], can be a significant source of deterministic error. This effect is generally more pronounced for high-quality devices that employ multiple, single axis sensors.

For the camera, we propose a direct formulation, motivated by the work of Meilland et al. [16]. This formulation can source more information per image than corner position based methods. It further circumvents the issue of assigning measurement timestamps at finer granularity than image exposure time: While start and duration of image exposure can be determined accurately, it is more difficult to resolve the time instant within the exposure window corresponding to a corner observation. More speculatively, the direct approach may further leverage the motion information comprised in motion blur, similar to the visual gyroscope proposed by Klein and Drummond [17], and it is able to treat defocus blur explicitly. The later incorporates insights from Joshi et al. [18] into estimating the point spread function for modelling blur. The approach differs from similar modelling proposed by Meilland and Comport [19] in that it estimates the blur kernel from data to achieve high-fidelity renderings that suffice the demanding requirements of calibration.

This work combines findings from our previous contribution on modelling accelerometer measurements as perceived in different locations inside the IMU [20] with a direct image measurement formulation [21]. It extends this work by a novel formulation of the direct image error that facilitates modelling uneven target illumination.

[1] The authors are with the ETH, the Swiss Federal Institute of Technology Zurich, Autonomous Systems Lab (www.asl.ethz.ch), Leonhardstrasse 21, LEE, CH-8092 Zurich, Switzerland. {joern.rehder@mavt.ethz.ch, rsiegwart@ethz.ch.

II. Method

A. Problem Statement

Most fundamentally, calibration aims at establishing a set of parameters $\boldsymbol{\Theta}$ that govern some sensor model $\mathbf{h}(\cdot)$ such that, given the system state $\mathbf{x}(t)$, $\mathbf{h}(\cdot)$ accurately predicts the measurement $\tilde{\mathbf{m}}$ for that sensor.

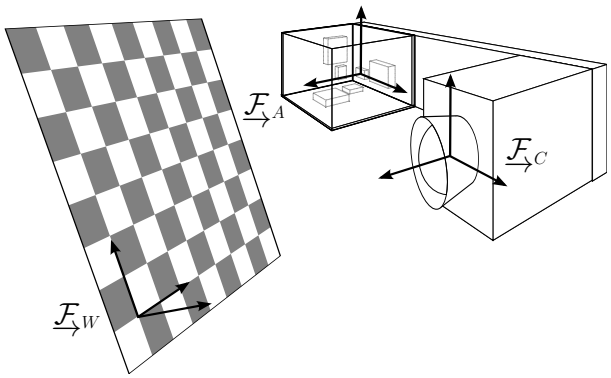


Figure 1: The general calibration setup. Camera and IMU are rigidly attached to each other and moved in front of a stationary visual calibration target. This work is concerned with finding the fixed transformation from the IMU reference frame, $\underrightarrow{\mathcal{F}}_A$, to the camera reference frame, $\underrightarrow{\mathcal{F}}_C$, as well as IMU intrinsics and a fixed temporal offset. Fig. 4 motivates these intrinsics by providing a close-up view into the internal structure of a prototypical IMU.

This work addresses the well-studied problem of camera/IMU calibration. Fig. 1 depicts the general calibration setup: A sensor suite, comprising an IMU rigidly attached to an intrsincially calibrated, global-shutter camera, is moved in front of a stationary visual calibration target. Using the images, accelerometer data, and gyroscope readings recorded in this process as measurements and an estimate of the sensor trajectory as state, a set of parameters comprising the relative pose between camera and IMU, a constant temporal offset and a set of intrinsic parameters of the IMU is determined.

Table I compiles all states and parameters estimated in this work.

B. Coordinate Frame Conventions

The different coordinate frames used in this work are shown in Fig. 1. We will refer to the target frame as $\underrightarrow{\mathcal{F}}_W$, while the camera

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2017.2674307, IEEE Sensors Journal

and the IMU frame will be denoted with $\underrightarrow{\mathcal{F}}_C$ and $\underrightarrow{\mathcal{F}}_A$ respectively.

The relative pose of two coordinate frames, e. g. of $\underrightarrow{\mathcal{F}}_W$ with respect to $\underrightarrow{\mathcal{F}}_C$, is fully described by means of a 4×4 transformation matrix $\mathbf{T}_{CW}$ denoting the mapping ${}_C\mathbf{p} = (\mathbf{T}_{CW})_W\mathbf{p}$ of points expressed in homogeneous coordinates:

$$\mathbf{T}_{CW} := \begin{bmatrix} \mathbf{R}_{CW} & {}_C\mathbf{t}_{CW} \\ \mathbf{0} & 1 \end{bmatrix}, \tag{1}$$

where $\mathbf{R}_{CW}$ is a rotation matrix and ${}_C\mathbf{t}_{CW}$ marks the vector from the origin of $\underrightarrow{\mathcal{F}}_W$ to the origin of $\underrightarrow{\mathcal{F}}_C$ expressed in $\underrightarrow{\mathcal{F}}_C$.

In addition to these coordinate frames, we will further consider the 2D image coordinate frame $\underrightarrow{\mathcal{F}}_I$ which describes coordinates on the image plane $\{{}_C\mathbf{p} : {}_Cz = 1\}$ scaled and translated according to the camera intrinsics $\mathbf{K}_{IC}$ introduced in (14).

Table I: States $\mathbf{x}(t)$ and parameters $\boldsymbol{\Theta}$ estimated in this work.

Symbol	Description	Section
	States $\mathbf{x}(t)$	
$\mathbf{T}_{AW}(t)$	Time-varying pose of the IMU	II-D
$\mathbf{b}_\alpha(t)$	Time-varying accelerometer bias	II-G1
$\mathbf{b}_\omega(t)$	Time-varying gyroscope bias	II-G2
	Parameters $\boldsymbol{\Theta}$	
$\mathbf{T}_{CA}$	Fixed relative transformation between $\underrightarrow{\mathcal{F}}_A$ and $\underrightarrow{\mathcal{F}}_C$	II-F1
d_C	Fixed temporal offset of image timestamps with respect to accelerometer timestamps	II-F1
$\mathbf{a}^k$	Coefficients of a polynomial illumination model for image k	II-F2
ρ_b^k	Reflectance of the black tiles of the calibration pattern for image k	II-F2
t_e	Exposure time of the camera	II-F2
o	offset of the linear camera response function	II-F2
$\mathbf{S}_{\alpha,\omega}$	Accelerometer and gyroscope scaling factors	II-G1,II-G2
$\mathbf{M}_{\alpha,\omega}$	Accelerometer and gyroscope misalignments	II-G1,II-G2
${}_A\mathbf{r}_{A\alpha_{y,z}}$	Displacement of accelerometer axes y and z from $\underrightarrow{\mathcal{F}}_A$	II-G1
$\mathbf{C}_{A\omega}$	Relative rotation between $\underrightarrow{\mathcal{F}}_A$ and the gyroscope frame $\underrightarrow{\mathcal{F}}_\omega$	II-G2
$\mathbf{A}_\omega$	Influence of linear acceleration on gyroscope measurements ("g-sensitivity")	II-G2
${}_W\mathbf{g}$	Direction of gravity expressed in $\underrightarrow{\mathcal{F}}_W$	II-G1

C. Estimator Formulation

The calibration is formulated as a Maximum Likelihood Estimation (MLE) over a batch of images and accelerometer and gyroscope measurements.

Each sensor contributes an error term of the form $\mathbf{e}_h := \mathbf{h}(\mathbf{x}(t), \boldsymbol{\Theta}) - \tilde{\mathbf{m}}$ to the estimator, where the measurement vector $\tilde{\mathbf{m}}$ is composed of all measurements recorded with the respective sensor and vector $\mathbf{h}(\cdot)$ comprises the corresponding, modelled values. We will further consider time-varying inertial sensor biases via process models of the form $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) + \mathbf{w}(t)$ where $\mathbf{w}(t)$ marks a zero-mean, white Gaussian process. This yields the corresponding contribution $\mathbf{e}_f(t) := \dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t))$.

We assume that all measurements $\tilde{\mathbf{m}}$ are corrupted by zero-mean, white Gaussian noise processes, either discrete in time as for the camera, or continuous in time for accelerometers and gyroscopes, the characteristics of which are captured by matrices $\mathbf{R}$. The processes $\mathbf{f}(\cdot)$ are modelled as affected by a zero-mean white Gaussian process with characteristics $\mathbf{Q}$.

With these assumptions, the estimator can be formulated as

$$\begin{aligned} \boldsymbol{\Theta}, \mathbf{x}(t) = \underset{\boldsymbol{\Theta}, \mathbf{x}(t)}{\operatorname{argmin}} \Big(& e_{h_C}(\mathbf{x}(t), \boldsymbol{\Theta})^T \mathbf{R}_C^{-1} e_{h_C}(\mathbf{x}(t), \boldsymbol{\Theta}) \\ & + e_{h_\alpha}(\mathbf{x}(t), \boldsymbol{\Theta})^T \mathbf{R}_\alpha^{-1} e_{h_\alpha}(\mathbf{x}(t), \boldsymbol{\Theta}) \\ & + \int \mathbf{e}_{f_\alpha}(\tau)^T \mathbf{Q}_\alpha^{-1} \mathbf{e}_{f_\alpha}(\tau) \mathrm{d}\tau \\ & + e_{h_\omega}(\mathbf{x}(t), \boldsymbol{\Theta})^T \mathbf{R}_\omega^{-1} e_{h_\omega}(\mathbf{x}(t), \boldsymbol{\Theta}) \\ & + \int \mathbf{e}_{f_\omega}(\tau)^T \mathbf{Q}_\omega^{-1} \mathbf{e}_{f_\omega}(\tau) \mathrm{d}\tau \Big), \end{aligned} \tag{2}$$

where subscripts C, α, and ω identify contributions as originating from the camera, accelerometer, and gyroscope model respectively.

We solve (2) iteratively for $\mathbf{x}(t)$ and $\boldsymbol{\Theta}$ using the Levenberg-Marquardt (LM) algorithm [22].

D. State Parametrization

Our implementation extends the open-source toolbox *kalibr* [23], which uses a continuous-time state parametrization. For completeness, we will present a brief introduction here that follows the original work very closely; for a detailed derivation of the underlying concepts, please see [24].

The state is represented as a weighted sum of a finite number of known analytical basis functions. Kalibr—and by extension this approach—employ B-splines as basis functions due to their simple analytical derivatives, good representational power and finite temporal support. The finite support yields a sparse system of equations in the estimator which can be solved efficiently.

A D-dimensional state, $\mathbf{x}(t)$, may be written as

$$\boldsymbol{\Phi}(t) := \begin{bmatrix} \boldsymbol{\phi}_1(t) & \dots & \boldsymbol{\phi}_B(t) \end{bmatrix}, \quad \mathbf{x}(t) := \boldsymbol{\Phi}(t)\mathbf{c}, \tag{3}$$

where each $\boldsymbol{\phi}_b(t)$ is a $D \times 1$ B-spline and $\boldsymbol{\Phi}(t)$ is a $D \times B$ stacked basis matrix. The state $\mathbf{x}(t)$ is then determined by estimating the $B \times 1$ coefficient vector $\mathbf{c}$.

The time-varying transformation $\mathbf{T}_{AW}(t)$ from $\underrightarrow{\mathcal{F}}_W$ into $\underrightarrow{\mathcal{F}}_A$ is parameterized as a 6×1 spline with 3 degrees of freedom for relative translation and 3 degrees of freedom for relative orientation:

$${}_W\mathbf{t}_{WA}(t) := \boldsymbol{\Phi}_t(t)\mathbf{c}_t \tag{4}$$

$$\boldsymbol{\varphi}(t) := \boldsymbol{\Phi}_\varphi(t)\mathbf{c}_\varphi\,. \tag{5}$$

In this work, we use the axis/angle parameterization for rotations, where $\boldsymbol{\varphi}(t)$ represents a rotation by the angle $\varphi = \sqrt{\boldsymbol{\varphi}(t)^T\boldsymbol{\varphi}(t)}$ about the axis $\boldsymbol{\varphi}(t)/\varphi(t)$. The orientation of $\underrightarrow{\mathcal{F}}_W$ with respect to $\underrightarrow{\mathcal{F}}_A$ at time t is given by

$$\mathbf{C}_{AW}(t) := \boldsymbol{\mathcal{C}}\big(\boldsymbol{\varphi}(t)\big)^T, \tag{6}$$

where $\boldsymbol{\mathcal{C}}(\cdot)$ is a function that builds a direction cosine matrix from the orientation parameters $\boldsymbol{\varphi}(t)$.

For both, orientation and translation, a sixth-order B-spline is employed, which encodes linear and angular acceleration as a cubic polynomial. The extent of the domain of support of individual basis functions is adjusted to match the expected bandwidth of the motion through the number of knots per second $N_\mathbf{x}$.

Time-varying sensor biases are represented by cubic B-splines

$$\mathbf{b}(t) := \boldsymbol{\Phi}_b(t)\mathbf{c}_b \tag{7}$$

with $N_\mathbf{b}$ knots per second.

E. Baseline Camera Measurement Model

The baseline approach [9] uses the projection of known 3D points ${}_W\mathbf{p}_m$ corresponding to corners in the visual calibration target to model camera measurements.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2017.2674307, IEEE Sensors Journal

The camera measurement function $\mathbf{h}_C(\cdot)$ is composed of contributions $\mathbf{h}_C^k(\cdot)$ from individual images $k \in [1, \ldots, K]$ as

$$\mathbf{h}_C\left(\mathbf{x}(t), \mathbf{\Theta}\right) := \begin{bmatrix} \mathbf{h}_C^1\left(\mathbf{x}(t), \mathbf{\Theta}\right) \\ \vdots \\ \mathbf{h}_C^K\left(\mathbf{x}(t), \mathbf{\Theta}\right) \end{bmatrix}, \tag{8}$$

where the $\mathbf{h}_C^k(\cdot)$ are calculated according to

$$\mathbf{h}_C^k\left(\mathbf{x}(t), \mathbf{\Theta}\right) := \begin{bmatrix} \boldsymbol{\pi}\left(\mathbf{T}_{CW}(t_k + d_C){}_W\mathbf{p}_1\right) \\ \vdots \\ \boldsymbol{\pi}\left(\mathbf{T}_{CW}(t_k + d_C){}_W\mathbf{p}_M\right) \end{bmatrix}. \tag{9}$$

Here, $\boldsymbol{\pi}(\cdot)$ denotes a projection function that maps from $\underrightarrow{\mathcal{F}}_C$ to $\underrightarrow{\mathcal{F}}_I$. The temporal offset d_C refers to a mismatch between the timestamp assigned to the image and the actual measurement instant, relative to the timing of the IMU. Without additional information, the source of the offset cannot be disambiguated [9].

The measurement vector $\tilde{\mathbf{m}}_C$ is constructed accordingly from the corresponding corner locations ${}_I\tilde{\mathbf{p}}_m^k$ in all k images with covariance $\mathbf{R}_C = \sigma_{{}_I\mathbf{p}}^2\mathbf{I}$, where $\mathbf{I}$ marks the identity matrix of matching size.

F. Direct Camera Measurement Model

This work further assesses a direct formulation of the camera model formulated on image intensities.

For this model, the contribution of a single image k to the camera measurement model $\mathbf{h}_C(\cdot)$ is given by

$$\mathbf{h}_C^k\left(\mathbf{x}(t), \mathbf{\Theta}\right) := \begin{bmatrix} \mathrm{B}\left({}_I\mathbf{p}_1^k, \mathbf{x}(t), \mathbf{\Theta}\right) \\ \vdots \\ \mathrm{B}\left({}_I\mathbf{p}_M^k, \mathbf{x}(t), \mathbf{\Theta}\right) \end{bmatrix}, \tag{10}$$

where $\mathrm{B}(\cdot)$ models image intensity, or brightness, at image points ${}_I\mathbf{p}_m^k$. For efficiency reasons, only a subset of all image points is used as detailed on in Section II-F3. The contribution to the measurement vector $\tilde{\mathbf{m}}$ is compiled analogously from the intensity values at ${}_I\mathbf{p}_m^k$ in image k. The noise process covariance is computed as $\mathbf{R}_C = \sigma_{\mathrm{B}}^2\mathbf{I}$.

The measurement model describes a mapping from the radiance $\mathrm{L}(\cdot)$ at some location ${}_W\mathbf{p}$ on the target onto image intensity $\mathrm{B}(\cdot)$ in the corresponding pixel location ${}_I\mathbf{p}$:

$$\mathrm{L}({}_W\mathbf{p}) \mapsto \mathrm{B}({}_I\mathbf{p}) \tag{11}$$

This mapping can be decomposed into a geometric component, the mapping from ${}_W\mathbf{p}$ to ${}_I\mathbf{p}$, and a radiometric one, the mapping from $\mathrm{L}(\cdot)$ to $\mathrm{B}(\cdot)$.

We use radiometric terms rather loosely in this work. Given that a single camera with unknown spectral response function is the sole source of information, it is impossible to obtain an estimate of the true sensor irradiance or of target illumination and reflectance. Instead, all estimates are distorted by the weighting of the spectral response curve and are only determined up to a scaling factor [25].

1) Geometric mapping: Rather than projecting a point ${}_W\mathbf{p}$ on the target onto coordinates ${}_I\mathbf{p}$ in the image, our approach performs the reciprocal mapping from ${}_I\mathbf{p}$ onto ${}_W\mathbf{p}$.

Assuming that the target is planar and aligned with the plane ${}_Wz = 1$, there exists a homography, $\mathbf{H}_{IW}^{-1}$, that maps from $\underrightarrow{\mathcal{F}}_I$ to $\underrightarrow{\mathcal{F}}_W$. For points $\{{}_W\mathbf{p} : {}_Wz = 0\}$ on the target, the mapping is computed as

$${}_I\mathbf{p} = \mathbf{H}_{IW}^{-1}\begin{bmatrix} {}_Wx \\ {}_Wy \\ 1 \end{bmatrix} \tag{12}$$

with homography

$$\mathbf{H}_{IW} = \mathbf{K}_{IC}\begin{bmatrix} {}^1\mathbf{R}_{CW} & {}^2\mathbf{R}_{CW} & {}_C\mathbf{t}_{CW} \end{bmatrix}, \tag{13}$$

where $\mathbf{K}_{IC}$ denotes the camera matrix, $\mathbf{R}_{CW}$ and ${}_C\mathbf{t}_{CW}$ are defined according to (1) and superscripts denote individual columns of the rotation matrix $\mathbf{R}_{CW}$.

Here, we will assume that $\boldsymbol{\pi}(\cdot)$ describes a pinhole camera model and that distortion has been compensated for. Other projection models are equally feasible, but require an adaptation of (13). For the pinhole model, camera intrinsics can be represented by the camera matrix $\mathbf{K}_{IC}$:

$$\mathbf{K}_{IC} := \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{14}$$

where $f_{x,y}$ denote the focal length and $c_{x,y}$ the principal point.

Casting the homography in terms of the sensor trajectory $\mathbf{T}_{AW}(t)$ and the fixed temporal offset d_C between the timestamp assigned to an image and the effective time period the sensor was exposed yields

$$\mathbf{H}_{IW}(t) = \mathbf{K}_{IC}\begin{bmatrix} \left(\mathbf{R}_{CA}{}^1\mathbf{R}_{AW}(t + d_C)\right)^T \\ \left(\mathbf{R}_{CA}{}^2\mathbf{R}_{AW}(t + d_C)\right)^T \\ \left({}_C\mathbf{t}_{CA} + \mathbf{R}_{CAA}\mathbf{t}_{AW}(t + d_C)\right)^T \end{bmatrix}^T \tag{15}$$

where $\mathbf{R}_{CA}$ and ${}_C\mathbf{t}_{CA}$ are the fixed rotation and translation relating $\underrightarrow{\mathcal{F}}_A$ to $\underrightarrow{\mathcal{F}}_C$.

2) Radiometric mapping: The radiometric part of the model is given by

$$\mathrm{L} \overset{\mathrm{S}(\cdot)}{\longmapsto} \mathrm{E} \overset{\int\int_A \mathrm{d}A\mathrm{d}t}{\longmapsto} \mathrm{X} \overset{\mathrm{R}(\cdot)}{\longmapsto} \mathrm{B} \tag{16}$$

where E and X mark sensor irradiance and exposure respectively and the functions $\mathrm{S}(\cdot)$ and $\mathrm{R}(\cdot)$ denote the optical transmission function and the sensor response function.

We will address all stages of (16) individually in the following.

The target radiance $\mathrm{L}(\cdot)$ is multiplicatively composed of the target's reflectance $\rho(\cdot)$ and an illumination term $\alpha(\cdot)$:

$$\mathrm{L}({}_W\mathbf{p}) = \rho({}_W\mathbf{p})\alpha({}_W\mathbf{p}) \tag{17}$$

For the checkerboard pattern, reflectance $\rho({}_W\mathbf{p})$ is given as

$$\rho({}_W\mathbf{p}) := \begin{cases} \rho_w & \text{if } (\lfloor \frac{{}_Wx}{\Delta x} \rfloor + \lfloor \frac{{}_Wy}{\Delta y} \rfloor) \bmod 2 = 0 \\ \rho_b & \text{else} \end{cases} \tag{18}$$

where the operator $\lfloor \cdot \rfloor$ denotes a floor operation and Δx and Δy the extent of individual checkerboard tiles in x and y direction. The values ρ_w and ρ_b mark the reflectance of the black and white tiles respectively. Since their true values and their ratio with respect to each other is unknown, we assume ρ_w to be 1, while ρ_b is estimated.

Target illumination is modelled as a 2nd degree polynomial:

$$\begin{aligned} \alpha({}_W\mathbf{p}) :=& a_1 + a_2{}_Wx + a_3{}_Wy + a_4{}_Wx^2 + \\ & a_5{}_Wy^2 + a_6{}_Wx{}_Wy, \end{aligned} \tag{19}$$

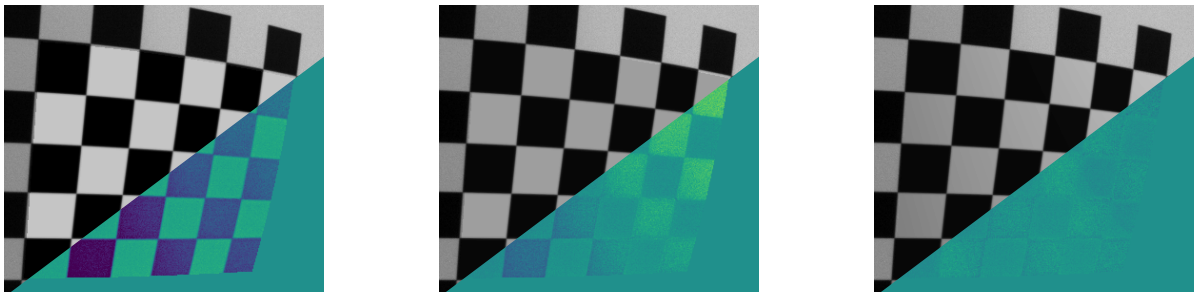
where $\mathbf{a} := [a_0, \ldots, a_6]$ denotes a set of model coefficients. This model is informed by the assumption that illumination varies smoothly over the target coordinates ${}_W\mathbf{p}$. Fig. 2 suggests that it has sufficient representational power to capture the nature of the lighting present in our experiments.

The sensor irradiance $\mathrm{E}(\cdot)$ results from applying the optical transmission function $\mathrm{S}(\cdot)$ to $\mathrm{L}(\cdot)$. This term commonly models vignetting [26]. The optics used in our experiments exhibit negligible dependence of attenuation on incidence angle. Accordingly, we assume constant attenuation and hence omit explicit modelling.

Sensor exposure $\mathrm{X}(\cdot)$ is defined as integral of the sensor irradiance over exposure time [25].

We further fold the integration over the finite extend of an individual pixel on the image sensor as well as the effect of imperfectly focussing optics into this step. Fig. 3a provides the rationale behind

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2017.2674307, IEEE Sensors Journal



(a) (b) (c)

Figure 2: The rendered target superimposed onto the corresponding camera image. The triangular insets show the color-coded difference image for clarity. Mismatches between image and model are most visible in two top rows. Fig. 2a shows the rendering prior to optimization. The intensity of the model clearly does not match the camera image. Fig. 2b displays the post optimization result using an even illumination model as proposed in [21]. The result exhibits subtle inconsistencies. Fig. 2c was generated using the polynomial illumination model (19). The image lacks any obvious visual seams, suggesting that a 2nd degree polynomial is sufficient to capture the lighting environment present in our datasets. Best viewed in color.

accounting for focussing effects: A perfectly focussed system would exhibit a sharp transition between checkerboard tiles, while real images show a more gradual change in intensities.

Correctly, exposure would be modelled as the integral of $g * \mathrm{E}({}_W\mathbf{p}(t))$ over the trajectory marked by ${}_W\mathbf{p}(t) = \mathbf{H}_{IW}^{-1}(t)\,{}_I\mathbf{p}$ during image exposure and over the area of the pixel, where the operator $*$ denotes a convolution and g marks the point spread function (PSF) of the optics.

We make the simplifying assumptions that the range of target depths present in the calibration dataset is sufficiently small such that the dependence of the PSF on distance can be neglected. We further omit its dependence on the position in image space [27].

We approximate the integrals as summations of the irradiance function discretized in time and image space. The convolution with the PSF is folded into the sum as discrete weights.

$$\mathrm{E}^*({}_I\mathbf{p}) := \frac{1}{J^2}\sum_{i=1}^{J}\sum_{j=1}^{J}\mathrm{W}_{ij}\mathrm{E}\left(\mathbf{H}_{IW}^{-1}(t)\left({}_I\mathbf{p} + \mu\begin{bmatrix}\frac{J}{i} - \frac{1}{2}\\ \frac{J}{j} - \frac{1}{2}\end{bmatrix}\right)\right) \tag{20}$$

$$\mathrm{X}({}_I\mathbf{p}) = \sum_{n=1}^{N}\mathrm{E}^*\left(t_0 + \frac{n}{N-1}t_e\right)\frac{1}{N}t_e \tag{21}$$

Here, N denotes the number of images rendered to emulate motion blur. Equation (20) marks a convolution of a super-resolution rendering of the irradiance image with a discretized kernel followed by down-sampling. In this view, J marks the size of the kernel, while μ denotes an up-scaling factor. The weights $\mathbf{W} := [W_{11}, W_{12}, \ldots, W_{JJ}]$ constituting this kernel are determined in a separate step using a set of static images of the calibration target. This calibration step is formulated as a minimization over the subset $\hat{\boldsymbol{\Theta}}$ of the parameters that govern the image forming process, excluding exposure time t_e which is unobservable for static images, as well as a set of static camera poses $\mathbf{T}_{CW}^k$ combined into state $\hat{\mathbf{x}}$:

$$\hat{\boldsymbol{\Theta}}, \hat{\mathbf{x}}, \hat{\mathbf{W}} = \underset{\boldsymbol{\Theta}, \mathbf{x}, \mathbf{W}}{\operatorname{argmin}} \sum_{k=1}^{K}\left(\mathbf{h}_C^k\left(\boldsymbol{\Theta}, \mathbf{x}, \mathbf{W}\right) - \tilde{\mathbf{m}}_C^k\right)^2 \tag{22}$$

Only positive weights are physically meaningful, which is enforced by estimating $\hat{W}_{ij} := \sqrt{W_{ij}}$ rather than W_{ij} directly. Without additional knowledge about illumination and reflectance of the target, the weights can only be determined up to an unknown scaling factor. Hence, we normalize the weights such that $\max(\mathbf{W}) = 1$.

For the optimization, $\hat{\boldsymbol{\Theta}}$ is initialized according to Section II-H and $\hat{\mathbf{x}}$ from target corners using the Perspective-n-Point algorithm [28], while all weights $\mathbf{W}$ are initialized to 1. Fig. 3b depicts the kernel estimated from 50 images for the optics used in the experiments in Section III.

The black square in the figure marks the boundaries of the respective pixel, highlighting that significant weights extend over an area of multiple pixels. Our current implementation lacks a principled approach to determining this extent and instead relies on multiple iterations of estimation (22) to determine a suitable combination of μ and J, starting from a large initial estimate for the kernel size J.

The sensor response $\mathrm{R}(\cdot)$ marks the mapping from sensor exposure to intensity.

Most sensors are designed for this mapping to be linear, and we disabled all digital processing of the signal in the sensor that would have altered the response curve. Accordingly, the camera response curve is modelled as linear as

$$\mathrm{B}({}_I\mathbf{p}) := s\mathrm{X}({}_I\mathbf{p}) + o, \tag{23}$$

where s denotes a scaling factor and o an offset. In this formulation, s is unobservable since any change could be compensated by scaling the illumination term (19) accordingly. Hence, s is assumed to be 1 and its estimation is omitted.

3) Camera error term reduction: Computing camera error terms is comparatively costly and not all pixels carry the same amount of information: Intensities at ${}_I\mathbf{p}$ corresponding to target locations ${}_W\mathbf{p}$ close to discontinuities in the reflectance function $\rho(\cdot)$ yield more information about the camera pose than points located centrally inside a checkerboard tile. Assuming that the initial estimate of the camera pose is sufficiently accurate, locations ${}_I\mathbf{p}$ with large gradients in the image $\tilde{\mathrm{B}}$ will correspond to informative locations on the target. This point selection is formalized as a classification function $\mathrm{C}(\cdot)$ depending on a gradient threshold τ where direct error terms are only evaluated for $\mathrm{C}({}_I\mathbf{p}) = 1$:

$$\mathrm{C}({}_I\mathbf{p}) := \begin{cases} 1 & \text{if } \left|\nabla\tilde{\mathrm{B}}_{{}_I\mathbf{p}}\right| > \tau \\ 0 & \text{else} \end{cases} \tag{24}$$

Despite this reduction, the resulting set of error terms will still yield a vastly over-constrained system of equations. Furthermore, the number of equations will change with the viewpoint.

We deem a large and varying number of error terms undesirable for implementation purposes and for reasons of computational efficiency. If the error terms were of fixed size, the block-sparsity pattern of the normal equation could be precomputed which would allow for more efficient solving [9].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2017.2674307, IEEE Sensors Journal

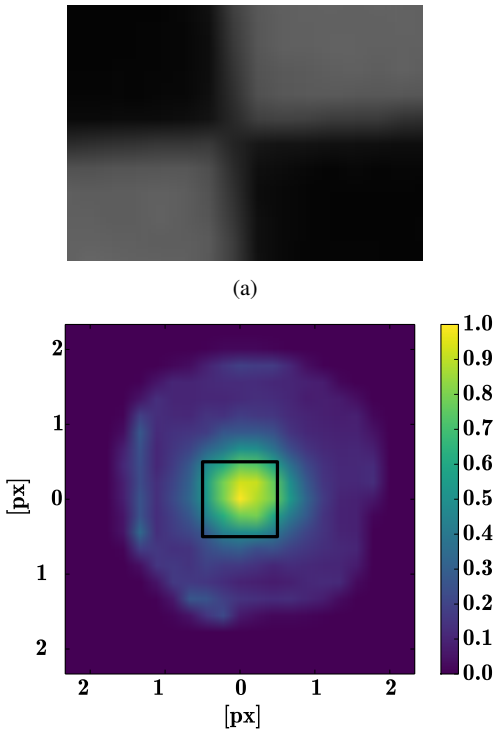


Figure 3: Imperfect focussing has a noticeable impact on image forming. Fig. 3a shows a magnification of a checkerboard corner recorded with our experimental setup at rest. For perfectly focusing optics, a narrow transition margin of 1 px between checkerboard tiles is expected. The real image exhibits a more gradual transition spanning multiple pixels. This behavior is modelled by rendering a super-resolution irradiance image, convolving it with the estimated blur kernel, Fig. 3b, and subsequently down-sampling the result.

Fixating the number of error terms is accomplished through QR-decompositions [29].

The Jacobian $\mathbf{J} := \left[\begin{array}{cc} \partial \mathbf{h}_C^k/\partial \boldsymbol{\Theta} & \partial \mathbf{h}_C^k/\partial \mathbf{x} \end{array}\right]$ of the camera measurement model associated with a single image k can be decomposed as

$$\mathbf{JP} = \mathbf{QR} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix} \tag{25}$$

where $\mathbf{P}$ is a column permutation matrix, $\mathbf{Q}$ is an orthogonal matrix, and $\mathbf{R}$ an upper triangular matrix. $\mathbf{P}$ is selected such that $\mathbf{R}_1$ is invertible by ensuring that all its diagonal elements are non-zero. The Jacobian of the reduced error term can be computed as

$$\hat{\mathbf{J}}^k := \mathbf{R}_1 \mathbf{P}^T \tag{26}$$

and the corresponding intensity error is given by

$$\hat{\mathbf{e}}_{h_C}^k := \mathbf{Q}_1^T \mathbf{e}_{h_C}^k. \tag{27}$$

G. IMU Measurement Model

The IMU model predicts accelerometer and gyroscope measurements given the sensor trajectory $\mathbf{T}_{AW}(t)$.

Accelerometers and gyroscopes contribute the terms $\mathbf{h}_\alpha(\cdot)$ and $\mathbf{h}_\omega(\cdot)$ to (2) as

$$\mathbf{h}_\alpha(\mathbf{x}(t), \boldsymbol{\Theta}) := \begin{bmatrix} \boldsymbol{\alpha}\left(\mathbf{x}(t_1), \boldsymbol{\Theta}\right) \\ \vdots \\ \boldsymbol{\alpha}\left(\mathbf{x}(t_K), \boldsymbol{\Theta}\right) \end{bmatrix} \tag{28}$$

and

$$\mathbf{h}_\omega(\mathbf{x}(t), \boldsymbol{\Theta}) := \begin{bmatrix} \boldsymbol{\varpi}\left(\mathbf{x}(t_1), \boldsymbol{\Theta}\right) \\ \vdots \\ \boldsymbol{\varpi}\left(\mathbf{x}(t_K), \boldsymbol{\Theta}\right) \end{bmatrix}, \tag{29}$$

with $\boldsymbol{\alpha}(\cdot)$ and $\boldsymbol{\varpi}(\cdot)$ as defined in (37) and (40) respectively, and where $t_k \in [t_1, \ldots, t_K]$ marks times at which the IMU recorded measurements. The contribution to the measurement vector is composed of accelerometer and gyroscope measurements $\tilde{\boldsymbol{\alpha}}_k$ and $\tilde{\boldsymbol{\omega}}_k$ accordingly, with corresponding noise covariance functions $\mathbf{R}_\alpha = \sigma_\alpha^2 \mathbf{I}\delta(t - t')$ and $\mathbf{R}_\omega = \sigma_\omega^2 \mathbf{I}\delta(t - t')$. Here, $\delta(\cdot)$ denotes Dirac's delta function, which is 1 for $t = t'$ and 0 otherwise.

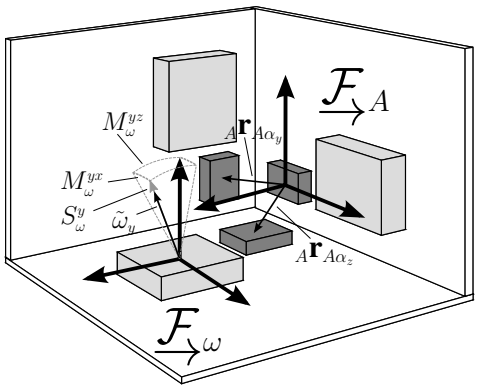


Figure 4: Conceptual drawing of the internal structure of an IMU composed of single-axis accelerometers (dark gray) and gyroscopes (light gray). We chose to align the input reference axes, $\underrightarrow{\mathcal{F}}_A$, with the accelerometer measuring in x direction. Consequently, the displacements ${}_A\mathbf{r}_{A\alpha_y}$ and ${}_A\mathbf{r}_{A\alpha_z}$ are estimated. Imperfections in the mechanical alignment yield both, non-orthogonal sensing axes, as illustrated by the misalignment terms M_ω^{yx} and M_ω^{yz}, and an unknown rotation between $\underrightarrow{\mathcal{F}}_A$ and $\underrightarrow{\mathcal{F}}_\omega$. Measurements might further be corrupted by an unknown scale factor S, visualized here as affecting $\tilde{\omega}_y$. These concepts equally transfer to IMUs realized inside a single integrated circuit (IC) despite their different mechanical design.

Fig. 4 shows the internal structure of an IMU schematically to illustrate the IMU intrinsics estimated in this work.

These intrinsics are an unknown rotation $\mathbf{C}_{\omega A}$ between $\underrightarrow{\mathcal{F}}_A$ and $\underrightarrow{\mathcal{F}}_\omega$, the displacements ${}_A\mathbf{r}_{A\alpha_{y,z}}$ of individual accelerometers with respect to $\underrightarrow{\mathcal{F}}_A$, misalignments of accelerometer and gyroscope axes with respect to the other axes as well as scale factor errors.

All IMU intrinsic parameters are further listed in Table I.

The inertial measurement models require linear acceleration, angular velocity, and angular acceleration which are derived from the continuous time formulation of the system state $\mathbf{x}(t)$ introduced in Section II-D.

Acceleration ${}_W\ddot{\mathbf{t}}_{WA}(t)$ is computed as

$${}_W\ddot{\mathbf{t}}_{WA}(t) = \ddot{\boldsymbol{\Phi}}_t(t)\mathbf{c}_t \tag{30}$$

from the spline parameters $\mathbf{c}_t$.

With $\mathbf{C}_{AW}(t)$ defined according to (6), angular velocity and angular acceleration as perceived in $\underrightarrow{\mathcal{F}}_A$ are computed as

$${}_A\boldsymbol{\omega}_{WA}(t) = \mathbf{C}_{AW}(t)\ {}_W\boldsymbol{\omega}_{WA}(t) \tag{31}$$

$${}_A\dot{\boldsymbol{\omega}}_{WA}(t) = \mathbf{C}_{AW}(t)\ {}_W\dot{\boldsymbol{\omega}}_{WA}(t) \tag{32}$$

with

$${}_W\boldsymbol{\omega}_{WA}(t) = \mathbf{S}\big(\boldsymbol{\varphi}(t)\big)\dot{\boldsymbol{\varphi}}(t) = \mathbf{S}\big(\boldsymbol{\Phi}(t)\mathbf{c}_\varphi\big)\dot{\boldsymbol{\Phi}}(t)\mathbf{c}_\varphi \tag{33}$$

$${}_W\dot{\boldsymbol{\omega}}_{WA}(t) = \mathbf{S}\big(\boldsymbol{\varphi}(t)\big)\ddot{\boldsymbol{\varphi}}(t) = \mathbf{S}\big(\boldsymbol{\Phi}(t)\mathbf{c}_\varphi\big)\ddot{\boldsymbol{\Phi}}(t)\mathbf{c}_\varphi \tag{34}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2017.2674307, IEEE Sensors Journal

where $\mathbf{S}(\cdot)$ is the matrix relating parameter rates to angular velocities and accelerations [30].

1) Accelerometer model: The specific force perceived by the accelerometers is composed of a component induced by the linear acceleration of $\underrightarrow{\mathcal{F}}_A$ relative to $\underrightarrow{\mathcal{F}}_W$, the gravitational force ${}_W\mathbf{g}$, and Euler and centrifugal forces induced by rotational motion *at the position of individual accelerometers*.

With $\boldsymbol{\omega}(t) := {}_A\boldsymbol{\omega}_{WA}(t), \dot{\boldsymbol{\omega}}(t) := {}_A\dot{\boldsymbol{\omega}}_{WA}(t)$, the specific force is computed as

$$\begin{aligned} {}_A\mathbf{a}_{WA}(t) = \mathbf{C}_{AW}(t)({}_W\ddot{\mathbf{t}}_{WA}(t) - {}_W\mathbf{g}) \\ + \operatorname{diag}(\lfloor\dot{\boldsymbol{\omega}}(t)\rfloor_\times \mathbf{R}_\alpha + \lfloor\boldsymbol{\omega}(t)\rfloor_\times^2 \mathbf{R}_\alpha), \end{aligned} \tag{35}$$

where $\operatorname{diag}(\cdot)$ extracts the $N\times 1$ vector from the diagonal of a matrix and operator $\lfloor\cdot\rfloor_\times$ denotes the skew-symmetric matrix that computes the cross product. The matrix $\mathbf{R}_\alpha$ is composed of the lever arms of individual accelerometers identified by subscripts according to

$$\mathbf{R}_\alpha := \begin{bmatrix} {}_A\mathbf{r}_{A\alpha_x} & {}_A\mathbf{r}_{A\alpha_y} & {}_A\mathbf{r}_{A\alpha_z} \end{bmatrix}. \tag{36}$$

We chose to align the position of the input reference axes (IRA) with the position of the x-axis accelerometer, i. e. ${}_A\mathbf{r}_{A\alpha_x} = \mathbf{0}$, and consequently do not include this quantity in the estimation.

Incorporating the IMU intrinsic parameters scaling, $\mathbf{S}_\alpha$, and misalignment, $\mathbf{M}_\alpha$, as well as a time-varying sensor bias $\mathbf{b}_\alpha(t)$, yields the complete accelerometer model

$$\boldsymbol{\alpha}(t) := \mathbf{S}_\alpha\mathbf{M}_{\alpha A}\mathbf{a}_{WA}(t) + \mathbf{b}_\alpha(t) \tag{37}$$

where $\mathbf{S}_\alpha$ is a diagonal matrix comprising scaling effects and $\mathbf{M}_\alpha$ is a lower unitriangular matrix, with off-diagonal elements corresponding to misalignment small angles.

The sensor bias $\mathbf{b}_\alpha(t)$ is modelled as being driven by a zero-mean, white Gaussian process [9]:

$$\dot{\mathbf{b}}_\alpha(t) = \mathbf{w}_\alpha(t) \tag{38}$$

with

$$\mathbf{w}_\alpha(t) \sim \mathcal{GP}\left(\mathbf{0}, \sigma_{b_\alpha}^2\mathbf{I}\delta(t-t')\right) \tag{39}$$

and hence $\mathbf{Q}_\alpha = \sigma_{b_\alpha}^2\mathbf{I}$.

2) Gyroscope model: Gyroscope measurements are modelled as

$$\begin{aligned} \boldsymbol{\varpi}(t) := \quad & \mathbf{S}_\omega\mathbf{M}_\omega\mathbf{C}_{\omega A}{}_A\boldsymbol{\omega}_{WA}(t) \\ & + \mathbf{A}_\omega\mathbf{C}_{\omega A}{}_A\mathbf{a}_{WA}(t) \\ & + \mathbf{b}_\omega(t) \end{aligned} \tag{40}$$

where $\mathbf{b}_\omega(t)$ marks the gyroscope bias. The rotation matrix $\mathbf{C}_{\omega A}$ denotes the unknown relative rotation between $\underrightarrow{\mathcal{F}}_A$ and $\underrightarrow{\mathcal{F}}_\omega$ and $\mathbf{S}_\omega$ and $\mathbf{M}_\omega$ are defined analogously to $\mathbf{S}_\alpha$ and $\mathbf{M}_\alpha$ in (37). The fully populated matrix $\mathbf{A}_\omega$ models the impact of the specific force on angular velocity measurements. Displacements of the gyroscopes from the IRA are not considered in ${}_A\mathbf{a}_{WA}(t)$, since the influence of the specific force on the measurement is insufficient to render these displacements properly observable.

The gyroscope bias is modelled analogously to (38) as

$$\dot{\mathbf{b}}_\omega(t) = \mathbf{w}_\omega(t) \tag{41}$$

with

$$\mathbf{w}_\omega(t) \sim \mathcal{GP}\left(\mathbf{0}, \sigma_{b_\omega}^2\mathbf{I}\delta(t-t')\right) \tag{42}$$

and $\mathbf{Q}_\omega = \sigma_{b_\omega}^2\mathbf{I}$.

H. Initialization

Sufficiently faithful initial estimates for the parameters $\boldsymbol{\Theta}$ and the state $\mathbf{x}(t)$ are required in order for (2) to converge to an accurate solution.

Most of the IMU intrinsic parameters are initialized assuming "perfect" sensors: The scaling factor matrices $\mathbf{S}_{\alpha,\omega}$ and the rotation between gyroscope and accelerometers $\mathbf{C}_{A\omega}$ are set to identity and misalignment $\mathbf{M}_{\alpha,\omega}$ and "g-sensitivity" $\mathbf{A}_\omega$ to $\mathbf{0}$. We initially assume that individual accelerometer axes perceive the specific force in an identical location, i. e. ${}_A\mathbf{r}_{A\alpha_{y,z}} = \mathbf{0}$.

The parameters governing the illumination model are initialized as 1 for coefficient a_1 and 0 for $a_{2,\ldots,6}$. Reflectance ρ_b and intensity offset o are initially set to 0. Exposure time t_e is initialized zero.

The estimates of $\mathbf{T}_{CA}$, the temporal offset d_C and the direction of gravity ${}_W\mathbf{g}$ are initialized from data. To this end, a set of camera poses $\hat{\mathbf{T}}_{CW}(t_k)$ for all image timestamps t_k is determined from corner observations by means of the Perspective-n-Point algorithm [28]. Subsequently, an orientation curve $\hat{\boldsymbol{\varphi}}(t)$, parametrized as a B-spline, is fitted to the camera orientations $\hat{\mathbf{C}}_{CW}(t_k)$. We employ a simplified model for the gyroscope measurements based on this orientation curve:

$$\hat{\boldsymbol{\varpi}}(t) := \mathbf{C}_{\omega A}\mathbf{C}_{CAC}^T\hat{\boldsymbol{\omega}}_{WC}(t) + \hat{\mathbf{b}}_\omega \tag{43}$$

The relative orientation $\mathbf{C}_{CA}$ and the constant bias $\hat{\mathbf{b}}_\omega$ are initialized to identity and zero respectively and subsequently estimated iteratively by minimizing

$$\mathbf{C}_{CA}, \hat{\mathbf{b}}_\omega = \underset{C_{CA},\hat{\mathbf{b}}_\omega}{\operatorname{argmin}} \sum_{k=1}^{K} \left(\hat{\boldsymbol{\varpi}}(t_k) - \tilde{\boldsymbol{\omega}}_k\right)^2. \tag{44}$$

The translation component of $\mathbf{T}_{CA}$, ${}_C\mathbf{t}_{CA}$, is initialized to zero.

Following the approach proposed by Mair et al. [7], the temporal offset d_C is initialized by correlating the absolute angular velocity as perceived independently by camera and gyroscopes. To this end, angular velocities ${}_W\hat{\boldsymbol{\omega}}_{WC\,k}$ are sampled from the spline $\hat{\boldsymbol{\varphi}}(t)$ at the timestamps t_k of gyroscope measurements. A coarse initial estimate for d_C is then derived as $d_C = d_{xcorr}T$ where T is the measurement interval of the gyroscopes and d_{xcorr} maximizes the cross-correlation between the two signals:

$$d_{xcorr} = \underset{d}{\operatorname{argmax}} \sum_{k=1}^{K} \left|{}_W\hat{\boldsymbol{\omega}}_{WC\,(k+d)}\right| \left|\tilde{\boldsymbol{\omega}}_k\right| \tag{45}$$

The direction of gravity ${}_W\mathbf{g}$ is initialized as the mean of the accelerometer readings transformed into $\underrightarrow{\mathcal{F}}_W$. Using the estimate of $\mathbf{C}_{CA}$ initialized with the previously introduced procedure and camera orientations $\hat{\mathbf{C}}_{CW}(t_k)$ sampled from $\hat{\boldsymbol{\varphi}}(t)$ at the time instants of the accelerometer readings, the initial value is computed as

$${}_W\bar{\mathbf{a}} = \frac{1}{K}\sum_{k=1}^{K} \hat{\mathbf{C}}_{CW}^{-1}(t_k + d_C)\mathbf{C}_{CA}\tilde{\mathbf{a}}_k \tag{46}$$

$${}_W\mathbf{g} = g_0\frac{{}_W\bar{\mathbf{a}}}{|{}_W\bar{\mathbf{a}}|}, \tag{47}$$

where g_0 is the magnitude of the gravitational acceleration.

Accelerometer and gyroscope biases are initialized to zero and the IMU trajectory is initialized by fitting a spline to the set of initial IMU poses, computed from camera poses $\mathbf{T}_{CW}(t_k)$ transformed by the initial estimate of $\mathbf{T}_{CA}^{-1}$.

III. Results

A. Experimental setup and dataset collection

All data were recorded with the visual/inertial sensor [31] shown in Fig. 5 featuring an Analog Devices ADIS16448 IMU, an InvenSense

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2017.2674307, IEEE Sensors Journal

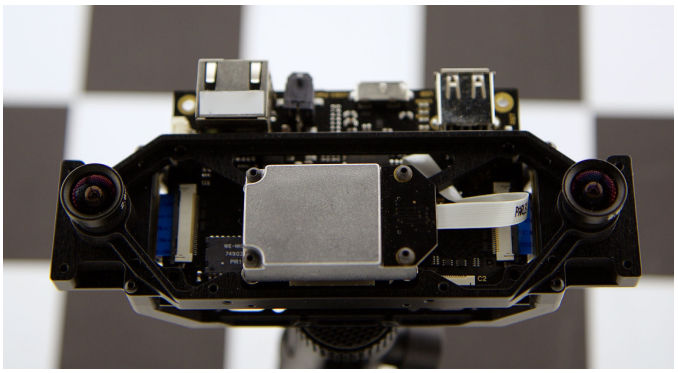


Figure 5: The experimental setup comprises an Analog Devices ADIS16448 IMU, an InvenSense MPU9150 IMU, and two Aptina MT9V034 global shutter image sensors of which only one was used.

MPU9150 IMU, and two Aptina WVGA MT9V034 global shutter image sensors of which only one was used. The ADIS16448 is a factory-calibrated MEMS device marketed specifically for navigation and robotics.

In contrast, the MPU9150 is a consumer-grade device. The camera was triggered at a rate of 20 Hz and used mid-exposure timestamping. IMUs were polled at 350 Hz. Timestamps for camera and IMUs were assigned by a single FPGA to avoid clock drift and limit jitter.

As calibration target, we used a checkerboard with square tiles of 70 mm. The board was illuminated by standard fluorescent office lighting.

We recorded 3 datasets by rapidly moving the sensor suite in front of a visual calibration target for about 200 s for each dataset. The datasets differed in camera settings: For the duration of each dataset, we fixated the exposure time to 0.96 ms, 2.24 ms, and 3.19 ms respectively. Indoor lighting conditions did not allow for exposure times significantly below 1 ms. The analog gain was further adjusted to yield similar image brightness across all datasets. We took care to excite all rotational degrees of freedom sufficiently without saturating the inertial sensors. Furthermore, we attempted to produce similar motion patterns for all datasets.

The approach exhibits a number of variables used to parametrize the algorithm as well as a number of noise parameters specific to the sensor setup. Table II lists all variables together with the values used to generate the results.

Table III compiles the noise model parameters. The noise model parameters for accelerometers and gyroscopes were determined using the approach proposed by Nikolic et al. [32]. The strength σ_{B} of the noise process acting on image intensities was determined from sequences of static images. The strength of the noise process assumed to affect the corner projections, $\sigma_{I\mathbf{p}}$, was determined from the preceding intrinsic calibration.

The direct approach is computationally significantly more expensive than the baseline method. On an Intel Core i7-2720QM at 2.2 GHz, the baseline method took on average about 30 s to converge on a 10 s chunk of data, while our implementation of the direct model required multiple minutes to find a solution.

B. Appropriate IMU modelling is key to high calibration precision.

The fidelity of the inertial measurement models directly impacts calibration performance.

Using the baseline approach (9) which models camera measurements as reprojection errors, this experiment asses the precision of camera/IMU extrinsics as well as of the time delay d_C. As input data 10 chunks of each 20 s length of the 0.96 ms dataset were used.

Table II: Variables used to parametrize the algorithm

Variable	Description	Value	Section
O	Order of the B-spline	6	II-D
$N_{\mathbf{x}}$	Knots per second supporting the pose spline	150	II-D
$N_{\mathbf{b}}$	Knots per second supporting the bias splines	50	II-D
τ	Threshold on the gradient in the image	7	II-F3
N	Number of images used to emulate motion blur	5	II-F2
J	Size of the weighting window $\mathbf{W}$	17	II-F2
μ	Up-scaling factor for rendering the irradiance image	3.5	II-F2

Table III: Noise model parameters

	Symbol	Value	Unit
Gyroscopes			
White noise str.	σ_{ω}	3.85×10^{1}	$^{\circ}\,\mathrm{h}^{-1}\,\sqrt{\mathrm{Hz}}^{-1}$
Bias diffusion	$\sigma_{b\omega}$	2.66×10^{-5}	$\mathrm{rad\,s}^{-2}\,\sqrt{\mathrm{Hz}}^{-1}$
Accelerometers			
White noise str.	σ_{α}	1.86×10^{-3}	$\mathrm{m\,s}^{-2}\,\sqrt{\mathrm{Hz}}^{-1}$
Bias diffusion	$\sigma_{b\alpha}$	4.33×10^{-4}	$\mathrm{m\,s}^{-3}\,\sqrt{\mathrm{Hz}}^{-1}$
Image sensor			
White noise str.	σ_{B}	1.98, 2.40, 1.77	—
White noise str.	$\sigma_{I\mathbf{p}}$	0.07	px

For both IMUs, three models of increasing complexity were considered. These models were

- assuming a perfectly calibrated IMU perceiving the specific force in a single spot, i. e. $\mathbf{S}_{\alpha,\omega} = \mathbf{I}$, $\mathbf{M}_{\alpha,\omega} = \mathbf{0}$, ${}_{A}\mathbf{r}_{A\alpha_{x,y,z}} = \mathbf{0}$, $\mathbf{C}_{\alpha,\omega} = \mathbf{I}$, and $\mathbf{A}_{\omega} = \mathbf{0}$.
- assuming an uncalibrated IMU perceiving the specific force in a single spot, i. e. ${}_{A}\mathbf{r}_{A\alpha_{x,y,z}} = \mathbf{0}$.
- assuming the full IMU model described in Section II-G.

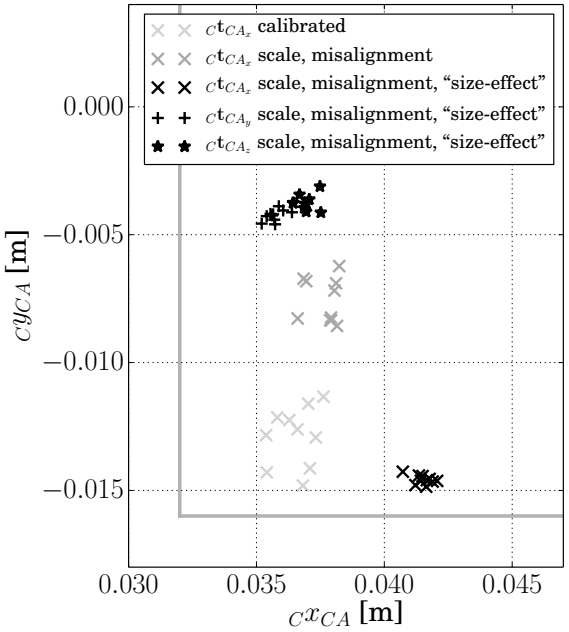
Table IV: Calibration results for the baseline camera error (9) and IMU models of different fidelity

IMU Model	$\boldsymbol{\sigma}_{C\mathbf{t}_{CA}}$ [mm]	σ_F [°]	σ_{d_C} [µs]
ADIS16448			
calibrated	$[0.75, 1.11, 0.52]$	0.040	19.05
uncalibrated	$[0.58, 0.83, 1.77]$	0.062	16.85
uncalibrated, size-effect	$[0.36, 0.17, 0.32]$	0.016	15.55
MPU9150			
calibrated	$[0.68, 1.02, 1.6]$	0.102	16.76
uncalibrated	$[0.11, 0.14, 0.16]$	0.008	1.92
uncalibrated, size-effect	$[0.15, 0.17, 0.25]$	0.008	2.13

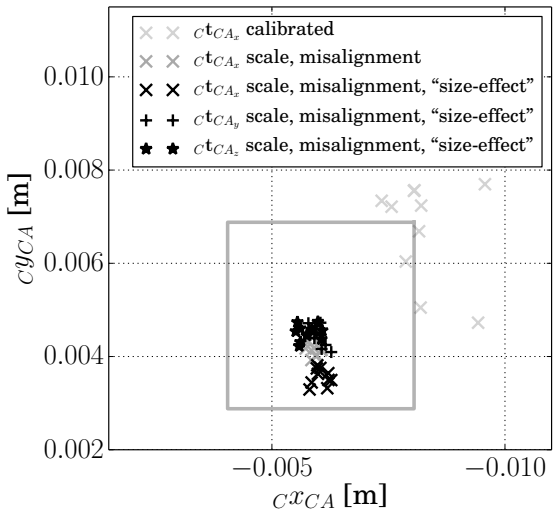
Fig. 6 depicts the ${}_{C}x_{CA}$ and ${}_{C}y_{CA}$ of the camera/IMU displacement ${}_{C}\mathbf{t}_{CA}$ as well as the position of individual accelerometer axes where estimated. Table IV displays the same results numerically.

Fig. 6a shows estimates for the ADIS16448, a factory calibrated, navigation-grade IMU. The *calibrated* and the *uncalibrated* model yield estimates of similar precision and located inside the IMU package which is highlighted as gray outline. Given that the device is factory calibrated, we would assume that scale factor errors and misalignments were compensated for by the manufacturer. Accordingly, calibration should not benefit from estimating these quantities. Table IV confirms this intuition, suggesting that including these parameters impacts the precision of extrinsic calibration negatively. This deterioration is consistent over the parameters relative displacement, relative orientation—assessed as the square root of the variance with

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2017.2674307, IEEE Sensors Journal



(a) ADIS16448



(b) MPU9150

Figure 6: Estimated displacement $_C\mathbf{t}_{CA}$ between camera and IMU for different levels of IMU model fidelity. The gray lines mark the approximate outline of the respective IMU packages measured in CAD drawings of the sensor setup.

respect to the Fréchet expectation [33] denoted as σ_F—and the temporal offset d_C alike. Fig. 6a further shows that estimating IMU intrinsics can result in a shift of the mean estimate of $_C\mathbf{t}_{CA}$. Including the estimation of the displacement of individual accelerometer axes into the calibration significantly increases precision of the parameters. It further reveals the presumed positions of the corresponding sensor elements as shown in Fig. 6a. While y and z axis are estimated to be in close vicinity to each other, the x axis element is displaced by about 1 cm, suggesting that it is housed in a different IC.

Fig. 6b shows results for the MPU9150, an uncalibrated, consumer-grade device. For this device, neglecting IMU intrinsics results in biased estimates located outside the package outline. Including intrinsic calibration yields improved calibration with significantly increased precision, where all estimates lie solidly inside the IMU package. These results confirm previous findings in literature ([10], [12], [20]) which suggested that neglecting IMU intrinsic calibration does not only decrease precision but also causes biased estimates. Calibrating for the size-effect deteriorates results again with increased standard deviations in the estimates of relative translation and orientation as well as d_C.

These findings are significant for a number of reasons: They show that the relative transformation between camera and IMU can be estimated to sub-millimeter precision and to below $\frac{1}{100}°$. They also confirm observations from [10] that the standard deviation in the estimates of d_C can be a small fraction of the measurement interval of the IMU.

The results further highlight that best calibration precision can be achieved for models that match the device: Estimating IMU intrinsics for calibrated units does not yield a benefit while it significantly improves results for uncalibrated devices. Conversely, determining the displacement of individual accelerometer axes boosts calibration performance for IMUs composed of multiple ICs while it slightly deteriorates precision in small devices.

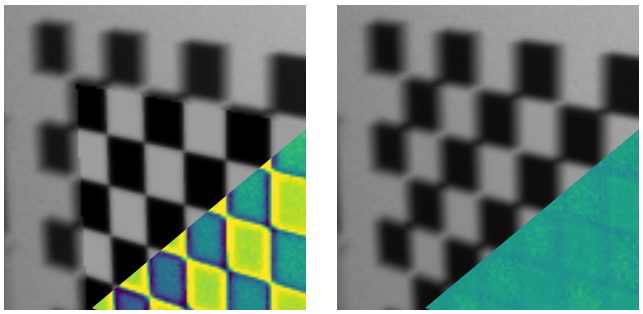
Given that this calibration approach shares much of the fabric of many visual/inertial state estimation frameworks, the results raise the question whether integrating a factory calibrated IMU pays off in all applications: The errors incurred by neglecting the displacements of individual accelerometer axes may devour all advantages of higher quality sensors and factory calibration—especially for applications with dominant motion patterns such as planar motion.

C. Exposure time can be accurately inferred from motion blur.

A prerequisite for the direct approach to yield accurate estimates is its capability to faithfully reproduce motion blur.

In this experiment, we used 10 segments of 10 s of each dataset.

Fig. 7 depicts the rendered target superimposed onto an image taken from the dataset at 3.19 ms exposure time. The exposure time is initialized to zero as introduced in Section II-H, resulting in the absence of motion blur in the rendered view shown in Fig. 7a. Following calibration, the exposure time is accurately estimated. Consequently, the rendered target closely resembles the image as apparent in Fig. 7b.



(a) (b)

Figure 7: Motion blur is emulated as an additive composition of target views rendered for a set of N camera poses spaced evenly over exposure time t_e. Fig. 7a depicts the rendered target superimposed onto an image prior to calibration. Fig. 7b shows the result after calibration, suggesting that the effect of motion blur can be accurately captured by the direct approach. Merely the absence of noise in the central patch of the checkerboard hints to its synthetic nature. Insets show color-coded difference images for clarity. Best viewed in color.

We use the estimated exposure time t_e as a proxy here to shed light on how accurately motion blur—and consequently the motion

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2017.2674307, IEEE Sensors Journal

of the camera during exposure—can be recovered. Fig. 8 shows a box plot of the exposure times t_e estimated for the 3 datasets. The narrow distribution of estimates close to their true value suggests that motion blur was equally accurately recovered. The mean and standard deviations of the exposure time estimates are 0.964 ± 0.007, 2.260 ± 0.013 and 3.21 ± 0.016 ms respectively.

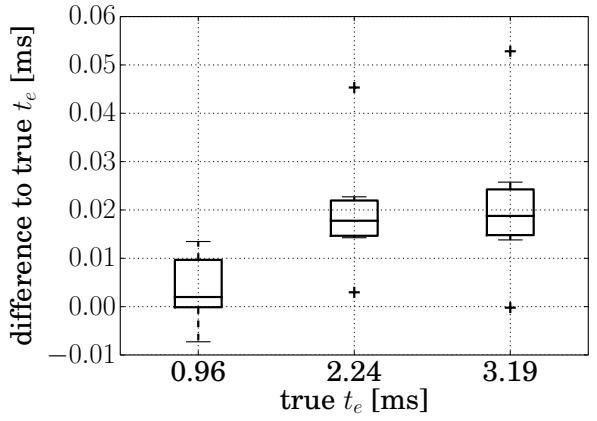


Figure 8: Estimation offset versus true exposure times for 3 datasets with different camera settings. The mean estimate is accurate to about $\frac{2}{100}$ ms, suggesting that the camera trajectory during exposure can be equally faithfully recovered.

These results are fundamentally different from our previous work [9], where exposure time was equally inferred from data. Conceptually, the previous approach estimated *half the exposure time* by consolidating information about the trajectory provided by the different sensor modalities by means of adjusting a fixed temporal offset. In contrast, this approach estimates exposure time by emulating motion blur in the images. Our experimental setup uses an exposure compensated triggering scheme as detailed on in [31] which renders t_e unobservable for our previous method.

D. The direct error formulation yields competitive results.

This experiment asses the direct camera measurement model on the same ten 20 s chunks used in Section III-B. It further exclusively focuses on the MPU9150 and the more sophisticated IMU models, since these combinations returned the most precise results. We fixated exposure time t_e to its nominal value of 0.96 ms and rendered images for 5 subsequent camera poses to emulate motion blur. Table V compiles the results achieved with these settings and assuming an uncalibrated IMU with negligible accelerometer displacements as well as an uncalibrated device and accounting for size-effect.

Table V: Calibration results for the direct camera error (10) and IMU models of different fidelity

IMU Model	$\boldsymbol{\sigma}_{C\mathbf{t}_{CA}}$ [mm]	σ_F [°]	σ_{d_C} [µs]
MPU9150			
uncalibrated	$[0.15, 0.18, 0.21]$	0.010	2.62
uncalibrated, size-effect	$[0.25, 0.20, 0.25]$	0.011	2.79

While precision is of a similar order as the results demonstrated in Section III-B, the direct approach performs slightly worse than the baseline.

Different reasons may contribute to this: First, our sensor suite employs a polling scheme for inertial measurements that retrieves data from the internal registers of the IMU at constant rate. The IMU itself sample internally at another constant rate. This scheme corrupts the timestamps of the measurements since the time of external polling—rather than internal sampling—is assigned as timestamp. The errors likely eclipse the improvements in timing resulting from the direct formulation. Furthermore, we noticed that the image sensor exhibits a number of isolated, "hot" pixels that behave differently from the rest of the sensor array and in turn cause large residuals. Such effects are currently not captured by the direct model and hence distort the result of optimization (2).

Nonetheless and despite the deteriorated performance of the direct method, the results suggest that modelling intensities rather than projections of corner points poses a viable approach to camera/IMU calibration. Improvements in the experimental setup as well as in modelling faulty sensor elements may enable the approach to leverage its presumed benefits detailed on in Section I.

IV. Conclusion

This work presented and assessed measures to increase precision in camera/IMU calibration. Improving the IMU model to account for the size-effect increased calibration precision significantly for our navigation-grade IMU. For the ADIS16448 IMU, the model clearly discerned the position of individual accelerometer axes. We saw similar separation of the x axis in the MPU9150, but were unable to equally clearly discern the location of the other two axes.

The direct formulation succeeded in accurately estimating exposure time, but failed to improve results over the baseline approach. We identified issues in the timestamping of inertial measurements in our experimental setup as well as a lack of modelling of "hot" pixels as potential sources of deteriorated performance. Future work will investigate these issues and extend the modelling of defocus effects to support different blur kernels in different parts of the image. We entertain the idea that motion blur may contain valuable information about the trajectory of the image sensor during exposure that may be leveraged to improve calibration, similar to the single frame visual gyroscope conceived by Klein and Drummond [17]. Future work may involve reproducing identical sensor trajectories for different exposure times to assess the value of this idea.

We believe that our results are the most precise reported to date for camera/IMU calibration (see [9], [10], [20], [34] for comparison). This precision can partly be attributed to improvements in the inertial measurement models. However, another key factor in increased precision over our previous work lies in a more dynamic calibration motion with average absolute angular velocities of around $270\,^\circ\,\mathrm{s}^{-1}$ as compared to about $150\,^\circ\,\mathrm{s}^{-1}$ in [20] and only about $55\,^\circ\,\mathrm{s}^{-1}$ in [9].

ACKNOWLEDGMENT

The authors would like to thank Janosch Nikolic and Amir Melzer for interesting discussions on the modelling of inertial sensors. We would further like to acknowledge everyone—but particularly Paul Furgale, Thomas Schneider, and Hannes Sommer—who contributed to the *kalibr* framework and whose efforts laid the groundwork for our implementation.

References

[1] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint kalman filter for vision-aided inertial navigation," in *Robotics and Automation (ICRA), 2007 IEEE International Conference on*. IEEE, 2007, pp. 3565–3572.

[2] E. S. Jones and S. Soatto, "Visual-inertial navigation, mapping and localization: A scalable real-time causal approach," *The International Journal of Robotics Research*, vol. 30, no. 4, pp. 407–430, 2011.

[3] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual–inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.

[4] J. Kelly and G. S. Sukhatme, "Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration," *The International Journal of Robotics Research*, vol. 30, no. 1, pp. 56–79, 2011.

[5] F. M. Mirzaei and S. I. Roumeliotis, "A kalman filter-based algorithm for imu-camera calibration: Observability analysis and performance evaluation," *IEEE transactions on robotics*, vol. 24, no. 5, pp. 1143–1156, 2008.

[6] M. Fleps, E. Mair, O. Ruepp, M. Suppa, and D. Burschka, "Optimization based imu camera calibration," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 3297–3304.

[7] E. Mair, M. Fleps, M. Suppa, and D. Burschka, "Spatio-temporal initialization for imu to camera registration," in *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on*. IEEE, 2011, pp. 557–564.

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

[9] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 1280–1286.

[10] J. Nikolic, M. Burri, I. Gilitschenski, J. Nieto, and R. Siegwart, "Non-parametric extrinsic and intrinsic calibration of visual-inertial sensor systems," *IEEE Sensors Journal*, vol. 16, no. 13, pp. 5433–5443, 2016.

[11] D. Zachariah and M. Jansson, "Joint calibration of an inertial measurement unit and coordinate transformation parameters using a monocular camera," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*. IEEE, 2010, pp. 1–7.

[12] C. Krebs, "Generic imu-camera calibration algorithm: Influence of imu-axis on each other," Autonomous Systems Lab, ETH Zurich, Tech. Rep., 2012. [Online]. Available: http://students.asl.ethz.ch/upl_pdf/396-report.pdf

[13] M. Li and A. I. Mourikis, "Online temporal calibration for camera–imu systems: Theory and algorithms," *The International Journal of Robotics Research*, vol. 33, no. 7, pp. 947–964, 2014.

[14] M. Li, H. Yu, X. Zheng, and A. I. Mourikis, "High-fidelity sensor modeling and self-calibration in vision-aided inertial navigation," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 409–416.

[15] J. Hung, J. Hunter, W. Stripling, and H. White, "Size effect on navigation using a strapdown IMU," U.S. Army Missile Research and Development Command, Guidance and Control Directorate Technology Laboratory, Tech. Rep., 1979.

[16] M. Meilland, T. Drummond, and A. I. Comport, "A unified rolling shutter and motion blur model for 3d visual registration," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2016–2023.

[17] G. S. Klein and T. Drummond, "A single-frame visual gyroscope." in *BMVC*, 2005.

[18] N. Joshi, R. Szeliski, and D. J. Kriegman, "Psf estimation using sharp edge prediction," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[19] M. Meilland and A. I. Comport, "Super-resolution 3d tracking and mapping," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 5717–5723.

[20] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart, "Extending kalibr: Calibrating the extrinsics of multiple imus and of individual axes," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*. IEEE, 2016, pp. 4304–4311.

[21] J. Rehder, J. Nikolic, T. Schneider, and R. Siegwart, "A direct formulation for camera calibration," in *Robotics and Automation (ICRA), submitted to the 2017 IEEE International Conference on*. IEEE, 2017.

[22] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. Springer, 2006.

[23] P. Furgale, J. Maye, J. Rehder, and T. Schneider. (2015, January) kalibr — a unified camera/imu calibration toolbox. [Online]. Available: https://github.com/ethz-asl/kalibr

[24] P. Furgale, C. H. Tong, T. D. Barfoot, and G. Sibley, "Continuous-time batch trajectory estimation using temporal basis functions," *The International Journal of Robotics Research*, p. 0278364915585860, 2015.

[25] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH 2008 classes*. ACM, 2008, p. 31.

[26] S. J. Kim and M. Pollefeys, "Robust radiometric calibration and vignetting correction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 4, pp. 562–576, 2008.

[27] F. Heide, M. Rouf, M. B. Hullin, B. Labitzke, W. Heidrich, and A. Kolb, "High-quality computational imaging through simple lenses," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 5, p. 149, 2013.

[28] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o (n) solution to the pnp problem," *International Journal of computer vision*, vol. 81, no. 2, pp. 155–166, 2009.

[29] G. H. Golub and C. F. van Loan, *Matrix Computations*. The John Hopkins University Press, 1996.

[30] P. C. Hughes, *Spacecraft Attitude Dynamics*. New York: John Wiley & Sons, 1986.

[31] J. Nikolic, J. Rehder, M. Burri, P. Gohl, S. Leutenegger, P. T. Furgale, and R. Siegwart, "A synchronized visual-inertial sensor system with fpga pre-processing for accurate real-time slam," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 431–437.

[32] J. Nikolic, P. Furgale, A. Melzer, and R. Siegwart, "Maximum likelihood identification of inertial sensor noise model parameters," *IEEE Sensors Journal*, vol. 16, no. 1, pp. 163–176, 2016.

[33] X. Pennec, "Probabilities and statistics on riemannian manifolds: Basic tools for geometric measurements." in *NSIP*. Citeseer, 1999, pp. 194–198.

[34] Z. Yang and S. Shen, "Monocular visual-inertial state estimation with online initialization and camera-imu extrinsic calibration," *IEEE Transactions on Automation Science and Engineering*, vol. PP, no. 99, pp. 1–13, 2016.

Joern Rehder is a Ph.D. student with the Autonomous Systems Lab at ETH Zurich. He received his M.Sc. in Electrical Engineering from the Hamburg University of Technology and has been a visiting scholar to the University of California, Berkeley, and the Field Robotics Center at Carnegie Mellon University. His current research is focused on design and calibration of visual/inertial sensor units.

Prof. Dr. Roland Siegwart received the M.Sc. and Ph.D. degrees in mechanical engineering from ETH Zurich, Zurich, Switzerland. He has been a Full Professor for Autonomous Systems with ETH Zurich, Zurich, Switzerland, since 2006 and the Vice President Research and Corporate Relations since 2010. From 1996 to 2006, he was an Associate and later a Full Professor for Autonomous Microsystems and Robotics with the Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland. He leads a research group of around 30 people working on several aspects of robotics. Dr. 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.